Naval Command,
Control and Ocean
Surveillance Center RDT&E Division San Diego, CA 92152-5001

Technical Report 1698
May 1995

Subjective Evaluation of Human-Computer Interface Options for a Tactical Decision Support System

Bernard K. Rummel
German Naval Medical Institute

Approved for public release; distribution is unlimited.

Technical Report 1698
May 1995

Subjective Evaluation of Human-Computer Interface Options for a Tactical Decision Support System

Bernard K. Rummel
German Naval Medical Institute

NAVAL COMMAND, CONTROL AND OCEAN SURVEILLANCE CENTER RDT&E DIVISION
San Diego, California 92152–5001

K. E. EVANS, CAPT, USN
Commanding Officer

R. T. SHEARER
Executive Director

ADMINISTRATIVE INFORMATION

The work detailed in this report was performed by the Naval Command, Control and Ocean Surveillance Center, RDT&E Division, Tactical Decision Making Under Stress Branch, Code 44216, for the Office of Naval Research. Funding was provided under program element 0602233N.

Released by
D. L. Conwell, Head
Deputy for Operations
Branch

Under authority of
J. D. Grossman, Head
Simulation and Human Systems
Technology Division

CONTENTS

Figures

Table

1. INTRODUCTION

1.1 BACKGROUND

The objective of the Tactical Decision Making Under Stress (TADMUS) Program is to apply recent developments in decision making theory and information display technology to the problem of enhancing tactical decision quality under stress. The technology will be demonstrated in the context of anti-air scenarios; general principles will be developed that will be applicable to other warfare areas. An experimental decision support system (DSS) will be produced with sufficient flexibility to permit examination of tactical decision making under conditions of stress. Researchers will evaluate the prototype in simulated tactical environments (Hutchins, 1995), initially in laboratory settings and later in operational settings. They will also evaluate display concepts for their ability to aid the decision maker(s) in acquiring and maintaining the ability to extract information rapidly and accurately from decision support systems under high stress loads.

This report documents results of a subjective evaluation study of display concepts for the experimental DSS. Prototype development was based on decision processes postulated by naturalistic decision making theory such as recognition-primed decision making (Klein, 1991, 1992 a, b), and explanation-based reasoning (Hair & Pickslay, 1992).

Initial considerations led to seven different display concepts. All concepts incorporated either the display principles suggested by naturalistic decision making theory or addressed typical errors observed in previous TADMUS experiments (Hutchins & Kowalski, 1993). Researchers drafted a human-computer interface (HCI) that presented the seven display concepts as independent DSS windows. Five of the seven windows could represent the underlying display principles in various ways, leading to different design options. Researchers referred to operational knowledge to find out which of these options was preferred. The NRaD TADMUS research team felt it was necessary to interview experienced fleet tactical action officers (TAOs) and commanding officers (COs) in order to address the following questions:

1. For windows where several design options exist, which option is the most preferred?
2. Why is this option preferred?
3. What modifications are recommended for the DSS HCI design options?
4. Which windows are considered to be useful, which are not?
5. How is the system likely to be used in tactical situations?

This report presents participants' preferences, interview and questionnaire data, and subjective usefulness ratings regarding the various display options investigated. The report's final section provides conclusions and recommendations on these results.

1.2 DESCRIPTION OF THE DSS HCI INVESTIGATED

1.2.1 Overall Description

The DSS supports a natural decision making process. The system presents raw and integrated data based previous research (Klein, 1992a, b; Kaempf et al., 1992) on human decision makers. The DSS does <u>not</u> make decisions of its own, but assists the user in certain stages of the decision making process. It supports "those processes which decision makers already use, rather than attempt to force decision makers to use some other strategy" (Smith & Grossman, in preparation).

The DSS design study investigated here included seven windows: Alerts, Track Profile, Comparison to Norms, Template, SABER, Response Manager Window, and Track Priority List. Researchers also investigated an additional rules of engagement (ROE) support function as part of the Response Manager Window. The following paragraphs describe each window's purpose.

1. Alerts Window: Provides a comparison of evidence of events to thresholds.
2. Track Profile Window: Presents the time history of a variable (altitude, range, speed) as an explicit feature.
3. Comparison to Norms Window: Provides a quick comparison of features for one contact to features for exemplars of other contacts.
4. Template Window: Assembles lower level features and compares them to reference values. Relates individual events, presents hypothesis for the situation based on integration of events.
5. SABER Window: Displays causal relationships between individual events, presents hypotheses for situation, based on causal model, presents evidence for hypothesis, presents assumptions required to accept hypothesis. It supports the Situation Assessment by Explanation-Based Reasoning process (Hair & Pickslay, 1992), e.g., the generation of a "story" explaining the current situation.
6. Response Manager Window: Provides assistance in using pre-planned responses.
7. Track Priority List Window: Provides an integrated picture that includes identification (ID), intent, priority and why, next action (pre-planned response), time to take next action.

Window Arrangement: Researchers arranged the seven windows on the display (figure 1) according to the following principles:

1. Activity and analysis windows were grouped separately.
2. Increasing level of integration and/or complexity were arranged from top to bottom.
3. Alerts Window and Track Priority List were central and thus placed in the left half of the screen, i.e., closest to the DEFTT Aegis screen.[1]
4. If user-made specifications in another window affected a window, researchers grouped both windows adjacently. Window arrangement allowed the user to maintain a continuous working cycle, preferably, clockwise.

1. Boff & Lincoln (1988) recommend placement of alerts messages not farther away than 30° from the display center). Furthermore, both windows used color coding to attract attention on important information. Since the sensitivity of the human eye to color information decreases beyond 50° off the fovea (Woodson, 1981), the color coding strategy was only useful close to the DEFTT screen.

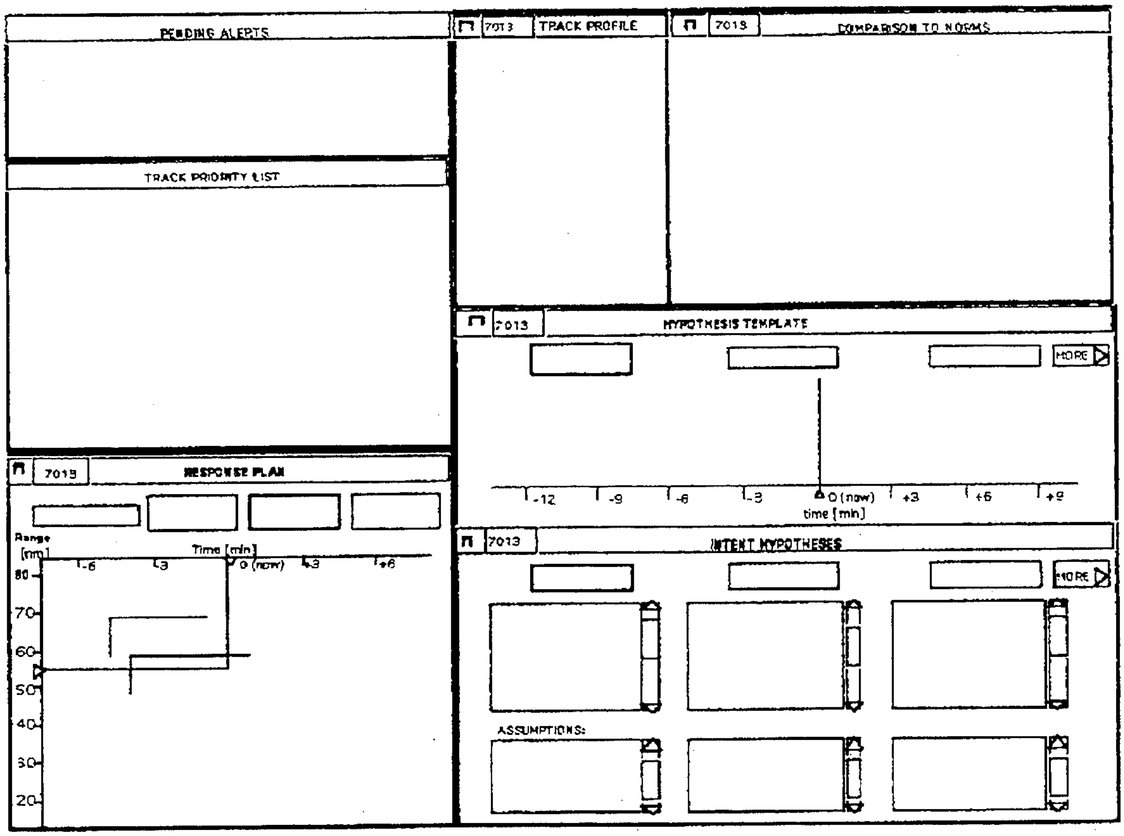

Figure 1. DDS Windows Display.

Some General Conventions. Researchers also used the following general conventions.

1. Displayed track information in white; system information, in light blue.
2. Displayed active screen elements, such as click-buttons and entry fields, in bluish gray rounded-corner rectangles.
3. Highlighted selected click-buttons.

1.2.2 Alerts Window

Researchers investigated two options for the Alerts Window. Both options followed a recommendation by Wickens (1984) on how to help an operator analyze abnormal situations. He suggested *sequencing* the single alerts. Wickens describes two strategies for presenting an alert sequence:

1. Flag single events of a sequence in the order of their occurrence, or,
2. Present the events themselves in a sequence.

The following options relate these two principles.

1. Option I (figure 2) consisted of a list of alerts, ordered by the time of occurrence. Increasing the track number ordered simultaneous alerts. The option displayed each alert message on a separate line with the corresponding track symbol, track number button, time of the alert, and a cancellation

PENDING ALERTS				
7013		**10:21**	**Turned inbound**	CANCEL
7020		10:20	Within Exocet range	CANCEL
7013		10:20	Within Exocet range	CANCEL
7013		10:18	Descending rapidly	CANCEL
7017	BOG	10:17	Within Exocet range	CANCEL
7013		10:15	Approaching keep out zone	CANCEL

Figure 2. Alerts Window (Option I, 75% of original size).

button. The alerts were color-coded corresponding to their importance (black on white text within red or yellow fields, or white text on blue background).

Researchers briefed participants that clicking on the cancellation button would eliminate the respective alert and that new alerts would flicker for 2 seconds to attract attention (cf. guidelines provided by Boff & Lincoln, 1988). Clicking on a track number button, or selecting a track for analysis in another window, caused highlighting of all buttons displaying the same track number (option I only). Bold characters simulated the flickering and highlighting so that a user could easily discern this track's alert messages. This corresponds to Wickens' (1984) approach of sequencing events by flagging.

2. Option II (figure 3): The option II display showed only the latest alert per track (the most recent alert still being in the top line). Researchers told the participants how to display a track's previous alerts. The participant clicks on a "History" button, then on a window popup button under the selected line. Checkmarks in this window designated previously canceled alerts. Clicking a "Return to list" button or specifying a different track for analysis returned the screen to the previous display. This corresponds to Wickens' (1984) approach of presenting a sequenced list of alerts.

PENDING ALERTS					
7013		**10:21**	**Turned inbound**	HISTORY▽	CANCEL
7020		10:20	Within Exocet range	HISTORY▽	CANCEL
7017	BOG	10:17	Within Exocet range		CANCEL

PENDING ALERTS				
7018	**10:23**	**Turned inbound**	HISTORY▽	CANCEL
7013	10:21	**Turned inbound**	**HISTORY▽**	CANCEL
RETURN TO LIST	10:20	Within Exocet range	√	
	10:18	Descending rapidly	√	
	10:15	Approaching keep out zone		

Figure 3. Alerts Window (Option II, 75% of original size). Lower display with history window active.

The history of a previously canceled alert was unavailable for display in the Option II Alerts Window since the alert line with the "History" button was no longer present. The alert history was available again only when a new alert occurred for the same track. Since there was only one alert per track displayed, and thus more room available, the cancellation button had a less important function in this display than in the option I display.

1.2.3 Track Profile Window

The Track Profile Window (figure 4) displayed time history graphs for select variables. Those variables included altitude (figure 4a), speed (figure 4b) and range (figure 4c) for air tracks, and speed and range for surface and subsurface tracks (figure 4d). Researchers told the participants how to select a variable by clicking on its button. Altitude was the default for air tracks, speed for surface and subsurface tracks. The range display included additional red, horizontal hairlines indicating the release ranges for typical weapons, e.g., Exocet, Harpoon, torpedo.

Researchers did not envision that actual data might adjust the different display scales. The altitude scale ranged from 0 – 40,000 feet, the speed scale from 0–1500 knots, and the range scale from 0 – 40 nautical miles. The time scale ranged from –12 minutes to 0 minutes in 3-minute steps to facilitate nautical calculations (a craft travels 1/10 of its speed [knots] in nautical miles in 6 minutes).

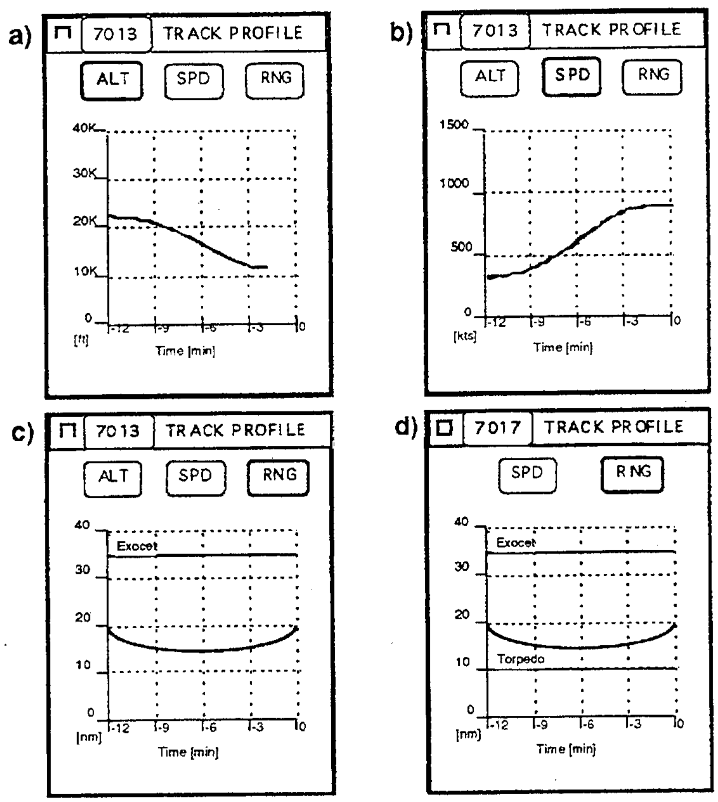

Figure 4. Track Profile Window for air (a–c) and surface (d) tracks (50% of original size).

1.2.4 Comparison to Norms Window

Researchers investigated three options for the Comparison to Norms Window.

1. Option I (figure 5): For a two-dimensional feature comparison task, a two-dimensional plot can display the critical variables and the respective ranges for different platform types. Points in a

two-dimensional altitude versus speed graph represented data from the contact of interest, the track symbol depicted the actual data; a line depicted the historical data. The display immediately illustrated if the track's data fit (or fit in the past) into a given platform type's range. The two-dimensional display allowed irregular and interdependent specifications for the speed and altitude ranges, e.g., a speed range varying with altitude. Colored lines surrounding range areas distinguished platform types of different threat levels (e.g., desaturated red for fighter aircraft).

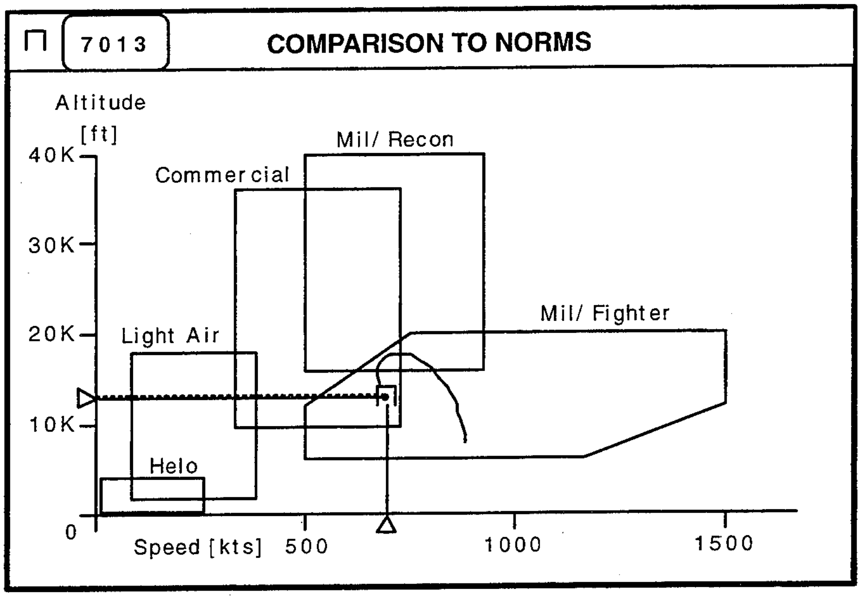

Figure 5. Comparison to Norms Window (Option I, 75% of original size).

This display type was "separable" in the terminology used by Bennett, Toms, and Woods (1993), since it allowed (raw) low-level data extraction by using the x and y axes as one-dimensional analog scales. However, the display was also "configurational" (in the same terminology) in the sense that it showed the "inside-ness" of data ranges as an "emergent feature" (Pomerantz, 1986).

2. Option II, and sub-options a, b, c (figure 6): The second option used discrete coding to provide information regarding whether a track's data fell within a platform type's specific data range or category. Three-level coding provides a quick comparison. The track's data either fit exactly into the respective range, or they are uncertain (e.g., fit within a certain deviation or are not interpretable), or they do not fit.

A two-dimensional matrix displayed these with the variables in the rows and the platform types in the columns. Researchers told participants how to click on the desired matrix cell to display exact data. The matrix columns' rounded corners indicated that researchers envisioned them as active screen elements.

Researchers examined three ways to code the data's fit: color coding (option IIa), fill pattern coding (option IIb), and cell shape coding (option IIc). In the color-coded version, desaturated green,

yellow, and red cell colors, respectively, indicate perfect data fit, uncertainty, and misfit. For coding by fill pattern, white (fit), scattered white or blue (uncertainty), and background color (misfit) were used as cell fill patterns. In the cell-shape, coded display, fill patterns were the same as in the fill-pattern, coded version, except for misfits that left the cell's space blank (background color only).

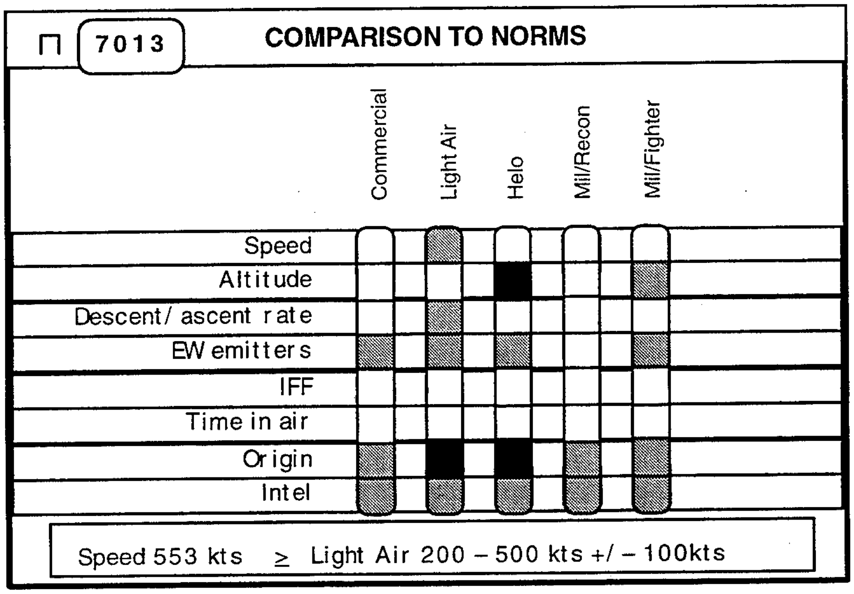

Figure 6. Comparison to Norms Window (Option II, 75% of original size).

The user's task was now, depending on the coding style, to look for the yellowish-green columns (IIa), or the columns without "holes" (IIb), or the "close-shaped" columns (IIc), to identify quickly in which column no misfits occurred. Researchers separated the matrix columns simplify this task.

This display was strictly "configurational" in the terminology of Bennett, Toms, and Woods (1993). The display's emergent feature (Pomerantz, 1986) was the integrity of columns indicated by common color, "filledness," or closure of shape. The display was not separable because researchers could not extract raw data from the matrix, although they envisioned their availability. Note that this display type provides an integration of many variables, and displays categorical data (such as electronic warfare (EW) emitter types).

3. <u>Option III</u> (figure 7): The option III window used analog scales to display speed, altitude, and descent or ascent rate. The window displayed the respective value ranges for different platform types as bars on the related scales. Each variable display's white line displayed the track's actual data. The variable's numerical value appeared under the scale title.

Researchers decided to include this display in the subjective evaluation study because of the findings by Sanderson, Haskell, and Flach (1992), who showed that "configurationality" does not necessarily improve performance. The team considered an analog scale the most basic display form for the comparison to norms task.

1.2.5 Template Window

Researchers envisioned the Template Window (figure 8) to display a track's particular behavior a specific intent, and to provide a comparison between a track's actual data and the template data.

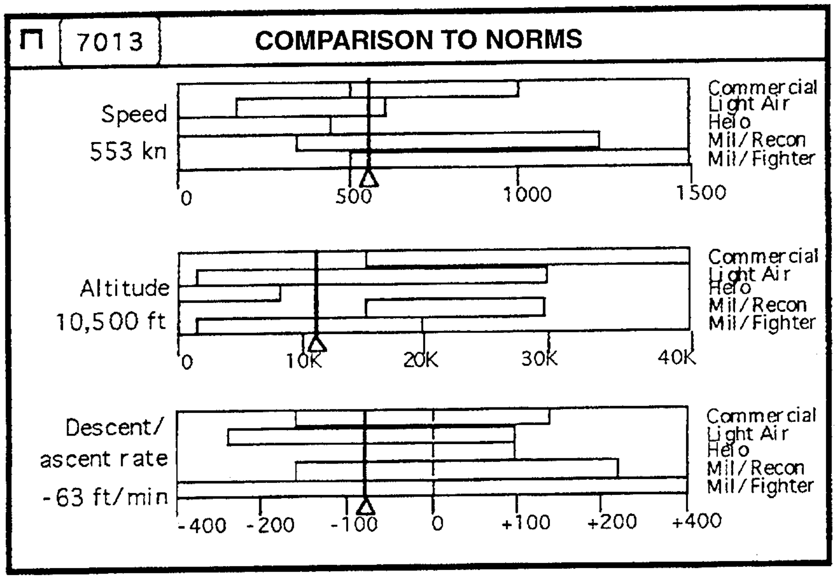

Figure 7. Comparison to Norms Window Option III (75% of original size).

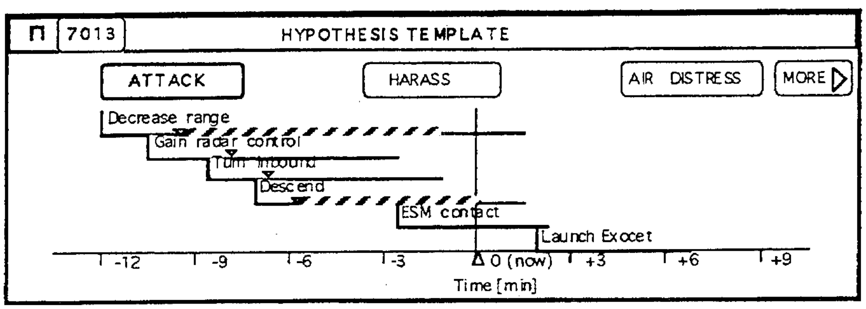

Figure 8. Template Window (50% of original size).

According to the intent hypothesis selected, the window displayed a track's expected behavior using bars representing the time range when the behavior occurred. The display superimposed the actual data when the behavior occurred, if it did. The window ordered the expected "behaviors" (e.g., "approaching," "descending," etc.) from top to bottom, following the common reading direction (cf. OSF/Motif™, 1990: Style Guide). The display indicated the respective time ranges as bars on a time scale (going from left to right), the bar's length representing the expected time range when the behavior will occur. Researchers envisioned this pattern to move along the time scale to the left (time "0," i.e., "now," indicated a stationery, vertical line). The display represented the actual data as triangular white markers for discrete events or the onset of a continuous event (like approaching), followed by diagonally hashed bars. The bars' length represented the event's duration.

Researchers told participants how to click on the respective buttons to select the display's underlying hypothesis from the Track Priority List, the SABER Window or the Template Window. The leftmost button (representing the most plausible hypothesis) was the default. If the displayed hypotheses' likelihood rank order changed, researchers envisioned the presentation order of the respective selection buttons to change accordingly. However, the selected hypothesis remained selected. If the user wanted to consider more than three hypotheses, researchers envisoned additional hypotheses.

Participants clicked on a "More" button on the right to display less likely hypotheses. They clicked on the left button for the three selection buttons' more likely hypotheses.

Researchers envisioned hypotheses disqualified by new evidence in the following way: for hypothesis not selected, the respective selection button would disappear. If a participant selected a hypothesis for display, a small message box appeared right under the selection button stating: "Disqualified by new evidence." When the participant selected another hypothesis or track, both the box and the button disappeared. The screen then displayed the three most likely hypotheses.

Researchers divided the time scale into 3-minute increments to simplify navigational calculations: the craft's speed in knots divided by ten equals the distance in nautical miles covered in 6 minutes.

1.2.6 SABER Window

The SABER Window (figure 9) displayed evidence and assumptions relating to three intent hypotheses at a time. The window displayed the most plausible hypothesis farthest to the left. Clicking the More button displayed more hypotheses. If the displayed hypotheses' likelihood rank changed, researchers envisoned changing the presentation order accordingly. They treated the hypothesis disqualified by new evidence the same way they treated it in the Template Window.

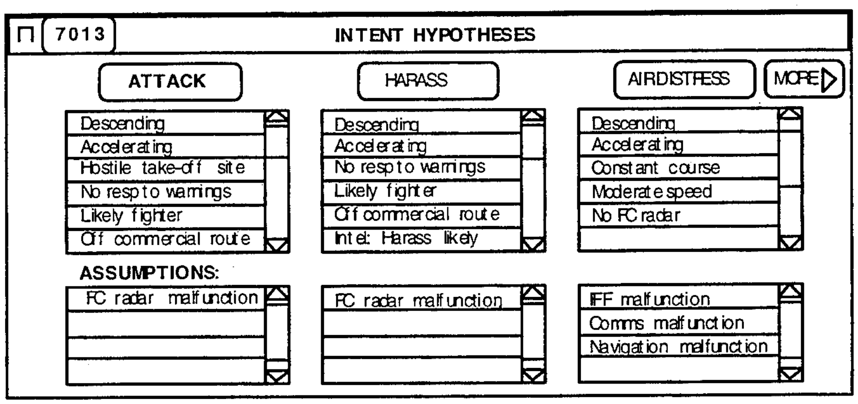

Figure 9. SABER Window (50% of original size).

Researchers told participants to use scrollbars to scroll evidence and assumptions lists. The scrollbars followed OSF/Motif™ standards (cf. OSF/Motif™, 1990: Style guide). The window displayed only supporting evidence for each hypothesis. The window did not display counter-indicators because researchers envisioned them to disqualify the hypothesis or to display supporting evidence under another hypothesis. However, the assumptions list provided any assumptions necessary for accepting the hypothesis.

Researchers anticipated that the evidence list would display several pieces of information, so they thought filtering the list could be useful. They considered the following options.

1. Option I: In the unfiltered list version, a user might have to scroll along the lists to look up the information of interest.

2. Option II: In the filtered list version, participants selected the evidence to differentiate among the three hypotheses. Thus, the window did not list evidence that was part of all three displayed hypotheses (e.g., "descending").

1.2.7 Response Manager Window

Researchers investigated two options for the Response Manager Window. In both options, researchers presented three general strategies, i.e., deconfliction, maximization of ownship safety and avoiding engagement, as select click buttons. Researchers envisioned the response plan to depend on the selected strategy.

1. Option I (figure 10): In option I, researchers arranged pre-planned actions on both a time and a range scale in a two-dimensional display. Researchers depicted the responses as time/range rectangles. They envisioned them moving to the left along the time scale. This indicated respective actions' time and range thresholds as well as the time delay permitted to perform them. The 0 = now line remained at the same place. Researchers envisioned the track's actual range as a horizontal line, moving downward as the track approached. To avoid overlap of the response rectangles, the window displayed only their upper and left border lines ("response angles"). Researchers told the participants that there was still time to perform a given response if the respective angle formed a closed rectangle with the 0 = now and the current range lines. A highlighted closed rectangle in the draft display informed the user what action the system recommended to take "right now."

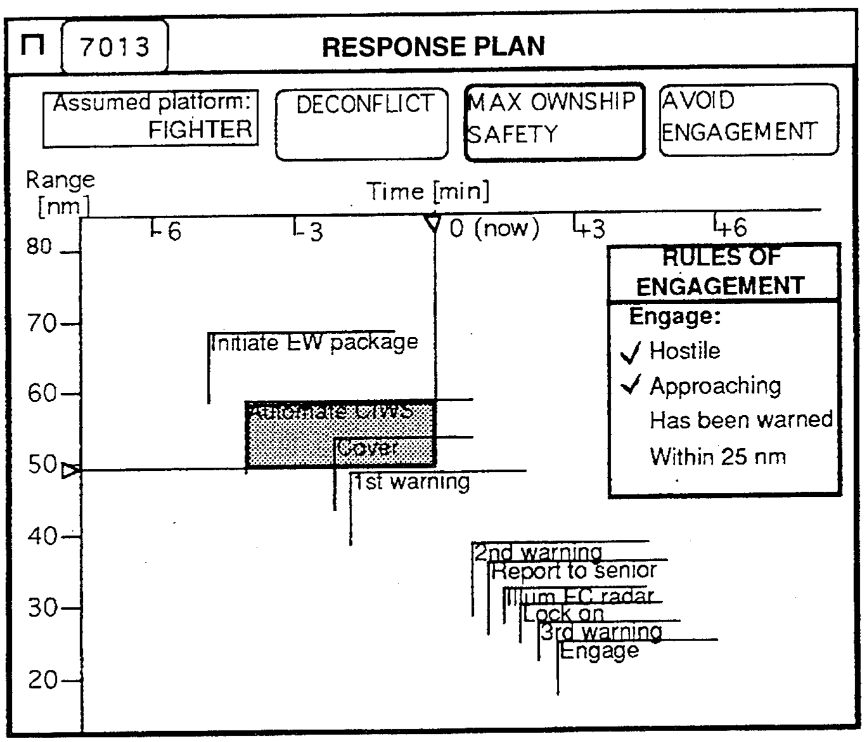

Figure 10. Response Manager Window with ROE box (Option Ia, 75% of original size).

Researchers anticipated that a given action's time/range angle size, as well as the angles' spacing on the horizontal axis, depended on the track's speed. This was because there is less time for action by range interval when the track is moving faster. The planned action's vertical spacing (= range spacing) would be unaffected by the track's speed.

2. Option II (figure 11): There was a certain redundancy in the option I display because the relation between the range and time axes caused a certain redundancy in the option I display. Option II took this into account by displaying only the time axis. In this option, researchers arranged the responses from top to bottom in the order performed. A time bar indicated each action. Researchers arranged the bar on the time axis to display the earliest and latest time to do an action. A table displayed the range thresholds. The action bars' equal vertical spacing in the option II display provided the opportunity to add prompt/feedback buttons to the window. Researchers told participants that when it was time to start an action, a button saying "Do it" appeared on the same line as the action bar. If the user clicked the button, the response "Done" appeared at the same spot. If the user did nothing and the time to take the action passed, the window highlighted the action specifier, and the button turned into a "Done?" button. The user had 1 minute to click this button to a "Done" statement. If the user still did not react within this time, the system assumed the user did not perform the and displayed the response "Missed."

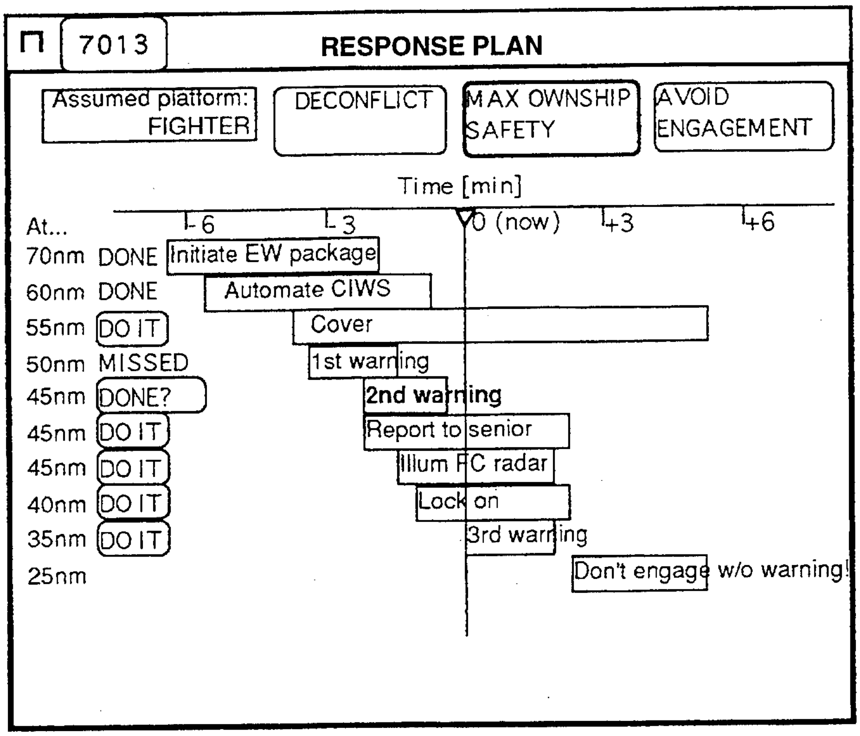

Figure 11. Response Manager Window with integrated ROE (Option Ib, 75% of original size).

This option might increase user workload because it is more interactive than the first option. However, the system required no other interface on actions.

1.2.8 Rules of Engagement Support

Preliminary results of the TADMUS experiments (Hutchins & Kowalski, 1993) indicated a high occurrence rate of failure to attend to Rules of Engagement (ROE). Researchers considered a ROE support function potentially helpful.

Researchers could display the ROE as a table inside the Response Manager Window (option a; see figure 10), or alternatively, integrate it in the display (option b; see figure 11). In the latter option,

e.g., the action "Engage" would not appear until it met the ROE criteria for engagement (i.e., warnings issued, track approaching, within 25 nmi, etc.).

Support for the ROE table comes from the view that rules of engagement do not relate immediately to the course of action decisions, but, highly important, they must be visible at *any* time. Increasing stress leads to decreasing working memory capacity (e.g., Hockey, 1986). Therefore, whether users choose to follow the recommended course of action in the Response Manager Window or not, they should always have access to the ROE.

Users tend to ignore a ROE table as a static display when they pay increasing attention to other parts of the display that may be continuously changing. This is especially likely under high stress conditions, when the user tends to focus on explicitly action-relevant cues. This view supports ROE integration in the recommended course of action. Integrating ROE in the course of action recommendations would ensure that users pay attention to them, but only if they consider the recommended course of action.

1.2.9 Track Priority List

Each line of the Track Priority List included the track symbol, track number button, up to three intent hypothesis buttons, the status of the track, the next action, and the permitted delay for this action. Researchers ordered lines by the operational priority assigned to the corresponding track.

To support a track's identification on the DEFTT geoplot, researchers could display the track's tag, if already assigned (option I; figure 12), or alternatively, they could display the track's bearing (option II; see figure 13). Note that users must enter a tag manually.

TRACK PRIORITY LIST.

Track	Assumed intent			Status	Next action	Perm. delay
7018.	Attk			**IMMED**	Engage	30s
7020	Attk	Har		**IMMED**	Lock on	30s
7013	**Attk**	Har	Recn	More Info	Illuminate	24s
7017 BOG	Attk	Har	Recn	More Info	Rep to senior	60s
7040	Attk	Har	Recn	More Info	Warn level1	60s
7014	Har	Recn	Attk	More Info	Initiate EW	120s
7019 LAC	Har	Recn	Attk	More Info.	Req crs chng	120s
7032	NOI			No Factor		
7011	NOI			No Factor		
7012	NOI			No Factor		
7001	NOI			No Factor		
7008.	NOI.			No Factor		

Figure 12. Track Priority List (Option I, 75% of original size).

TRACK PRIORITY LIST.

Track	Brg	Assumed intent			Status	Next action	Perm. delay
7018	153	Attk			**IMMED**	Engage	30s
7020	025	Attk	Har		**IMMED**	Lock on	30s
7013	275	**Attk**	Har	Recn	More Info	Illuminate	24s
7017	175	Attk	Har	Recn	More Info	Rep to senior	60s
7040	050	Attk	Har	Recn	More Info	Warn level 1	60s
7014	345	Har	Recn	Attk	More Info	Initiate EW	120s
7019	010	Har	Recn	Attk	More Info	Req crs chng	120s
7032	035	NCI			No Factor		
7011	150	NCI			No Factor		
7012	090	NCI			No Factor		
7001	210	NCI			No Factor		
7008	350	NCI			No Factor		

Figure 13. Track Priority List (Option II, 75% of original size).

2. METHOD

2.1 PARTICIPANTS

Seven Commanding Officers (COs) and nine Tactical Action Officers (TAOs) participated in the subjective evaluation study. Most had participated in a TADMUS experiment before (Hutchins & Kowalski, 1993). All were shipboard-qualified Navy officers (O 2 to O 6) with Aegis or New Threat Upgrade (NTU) experience. They had 5 to 30 years of service (average 18.3 years) and 6–100 months of experience standing CO or TAO watch (average 45.9 months) on one to six different Aegis/NTU ships (average 2.6). All had experience in littoral warfare with deployments in the Persian Gulf or the Mediterranean (average 2.3 PERGULF deployments; all after 1987, 75% after 1991). Table 1 in Appendix B gives a summary of the participants' demographic data.

2.2 PROCEDURE

Researchers collected data in the Decision Making Evaluation Facility for Tactical Teams (DEFTT; see Hutchins & Duffy, 1992 for a description) laboratory at NRaD. As a warm-up, participants watched a video tape of their performance in one TADMUS experimental scenario. Researchers encouraged the participants to comment on their actions in the scenario. The objective was to induce a state of self-awareness, and to establish a communication base to talk about thought processes involved in tactical decision making situations. A subject matter expert presented a briefing on TADMUS scenario B to officers who had not previously participated in a TADMUS experiment.

Researchers presented a slide show on the DSS and explained the system's purpose and function. They also explained the arrangement of the windows on the screen. Early in the scenario, the researchers displayed the video sequence in "freeze" mode to provide a visual cue regarding the operational situation.

Researchers presented and discussed all DSS windows, and presented alternative options. They asked participants to indicate their preferences and recommendations for the various options for the Alerts, Comparison to Norms, SABER, and Response Manager Windows, and the Track Priority List. A questionnaire (Argument Importance Rating Questionnaire) that listed arguments for every option supported the interview. This helped determine the relative weight of pro and con arguments. Researchers added a participant's new arguments to the questionnaire for the next participant. Participants checked symbols to indicate whether they considered an argument crucial, of some importance, or negligible. For arguments that were not crucial (i.e., rendering any other argument obsolete), participants chose symbols to provide interval scale properties. Figure 14 depicts the symbols used.

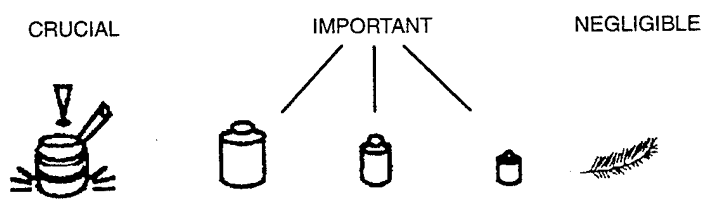

Figure 14. Symbols used in the Argument Importance Rating Questionnaire

Researchers asked participants to indicate their opinions regarding the windows, i.e., which ones they considered very useful, useful, or useless. A file card sorting procedure supported this part of the session. Participants sorted cards with the windows' names (including all options) into three piles

indicating the usefulness of the windows. Researchers also encouraged them to tell how they would use the windows tactically.

At the end of the procedure, researchers built a screen using a participant's preferred options. The participant watched a running TADMUS scenario. Researchers then asked the participant how they would use the DSS in the running scenario.

Throughout the session, researchers encouraged participants to make any comments or recommendations they had about a window, and to ask questions. They recorded the participant's responses as anecdotes. The "Procedure Script" (see Appendix A) describes the procedure in more detail, including the instructions given to participants.

2.3 DATA ANALYSIS

There were four categories of data (dependent variables):

1. Preferences for window options. Researchers tested the preference orders for window options for statistical significance using the sign test (if considering two options) or the Friedman two-way variance analysis (if considering more options). Using correlation data, researchers attempted to group participants into "opinion clusters." Thus, they examined the sample's homogeneity. Researchers expected that COs and TAOs would form separate groups.

2. Questionnaire data. Researchers conducted Friedman analyses of rank variance to evaluate the relative differences of argument weights. They also used Kendall's coefficient of concordance to estimate agreement between participants.

3. Three-category window usefulness ratings. Researchers used a Friedman two-way analysis of variance to determine inter-rater agreement. Using correlation data, they attempted to group participants into "opinion clusters." Thus, researchers examined the sample's homogeneity. They expected that COs and TAOs would form separate groups.

4. Free-format comments and interview data. Researchers collected interview data by tape recorder and as anecdotal interview protocols on paper. According to the heuristic approach chosen, researchers analyzed interview data as lessons learned from the research questions formulated in section 1. They described the participants' personal styles as they emerged from the interview data, and tentatively assigned participants to style prototypes.

3. RESULTS

This report presents results by windows, not by data types. The following paragraphs present an introduction on how to interpret the results.

1. Interview data reported in prose. Appendix B lists the original notes. References to a participant's statement are in brackets, such as "(5)." This identifies personal opinions (see table 1 in Appendix B for demographic and background information). Researchers grouped participants' according to the research questions investigated, under the categories "recommendations for experimental HCI," "factors influencing preference/subjective usefulness," "use of the display," and "future HCI."

2. Questionnaire data: Researchers coded questionnaire items as follows. AL1M2, for example, is the second argument in favor of the Alerts Window option I. TPL2M3 indicates the third argument outlining a disadvantage of the Track Priority List option II. The symbols used in the questionnaire appear in the respective summarizing figures and in figure 14 to simplify interpretation. The number of observations varied between individual questionnaire items because researchers introduced some items after the first interview sessions. They dropped some of those items for statistical analyses because of an insufficient number of observations.

3.1 GENERAL RESULTS

3.1.1 Clusters and Subgroups Among Participants

Researchers did not find meaningful opinion clusters for preference or subjective usefulness data through statistical cluster analyses. A factor analysis on preference data yielded a three-factor solution representing 75% of the variance in which researchers randomly distributed participants' data. However, a different approach successfully explained this finding.

Researchers calculated separate factor analyses for both preference and subjective usefulness data, using the transposed data matrices. Thus, they grouped participants rather than variables into factors. Using Pearson correlation coefficients and the Kaiser criterion (eigenvalues >1) in both cases, researchers found two-factor solutions that explained 92.8% of the preference variance and 86.8% of the subjective usefulness variance.[2]

These results allowed mapping participants into a two-dimensional space using a Multidimensional Scaling approach. Figure 15 shows such a map based on preference data.[3] A fairly linear Shepard diagram and a final Kruskal stress value of .05 indicated that the mapping approach was correct in this case. Note that such maps do not have necessarily interpretable axes, but show proximity relationships between objects (participants). COs form a fairly close cluster in the present map (participants 8, 11, 12, 14, 15, 16), while TAOs do not. This explains why statistical cluster analyses did not yield interpretable results. Researchers also interpreted this analysis as a caveat regarding comparisons between COs and TAOs, because TAOs did not form a distinct group. In view of the variance inhomogeneities between the CO and TAO groups apparent in figure 15, tests for average differences are not meaningful.

2. Note that the Pearson correlation coefficient tends to reduce the number of factors in the current case, assuming continuous distribution of values; analyses using Kendall's Tau detected more factors explaining less variance.

3. The corresponding map using subjective usefulness data shows similar results.

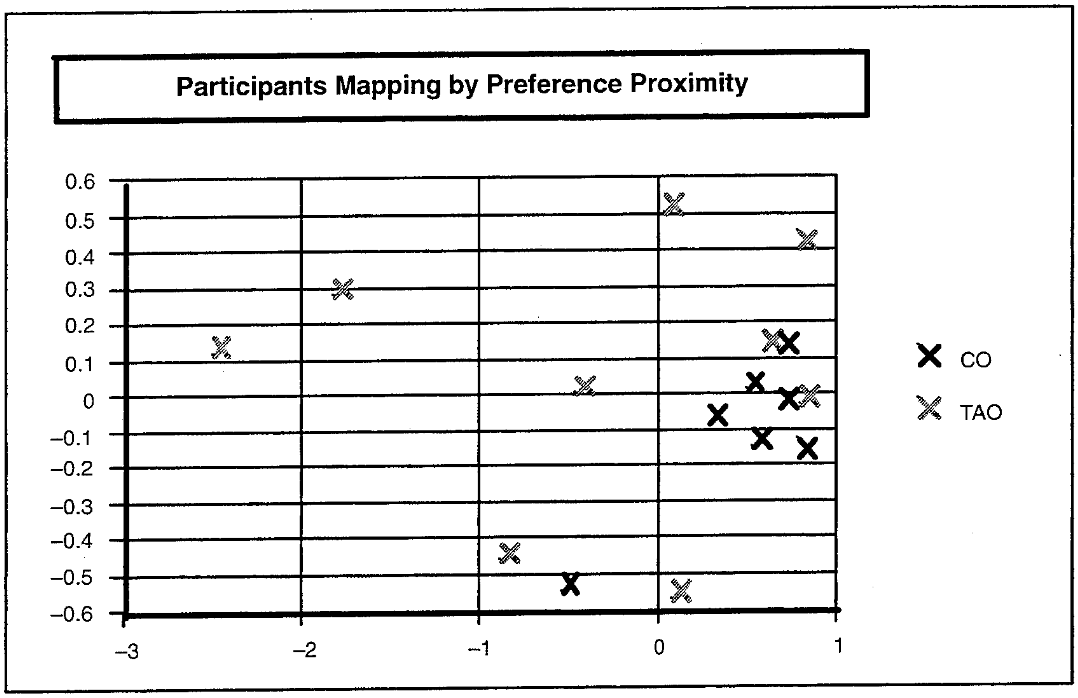

Figure 15. Two-dimensional map of participants using Pearson correlation coefficients of preference data.

Researchers analyzed CO cluster references separately. This group of participants exhibited the same preference pattern as the overall sample, but with more agreement within the group. Therefore, this report does not discuss CO cluster preferences separately unless there are marked differences.

3.1.2 Usefulness Ratings

Researchers used a sign test to see if all individual window options' subjective usefulness was greater than zero. Researchers considered all options except the Comparison to Norms Window (option III) as useful at an amount significantly different from zero ($p \leq .016$; for Comparison to Norms Window option III, $p = .125$).

3.1.3 Interview Data Prior to DSS Presentation

The following section presents comments made by participants before viewing the TADMUS scenario.

1. Importance of determining identification: One participant (14) stressed that the integration of information to determine intent and identification (ID) is crucial. He considered this the "weak link in current systems," and felt the "key watch station is the one that works on ID." However, the CO would not work on the ID problem, but would have to "accept ID as given and think what to do about it."

2. Important track features: Several participants indicated features and variables that are important to them. However, there was no common variable set. Most participants agreed on the importance of the altitude and kinematic profile (14, 10, 13). However, several indicators were subject to individual preferences: range (10); being within launching range (13), on a commercial route (10), or inbound (10, 13); responding to warnings (13). However, one CO (11) would only accept visual identification, non-cooperating target recognition (NCTR), and lack of mode 4 Identification Friend or Foe (IFF), if all were detected by the same aircraft or section.

Integration: A CO (16) strongly recommended integration of the DSS and the geoplot: a "user must be able to hook on a track for analysis from the geoplot. Sometimes you want to analyze a track that is not yet on the Track Priority List. The CO will not enter track numbers "manually." This suggestion was made several times in different contexts.

3.2 RESULTS FOR INDIVIDUAL WINDOWS

3.2.1 Alerts Window

3.2.1.1 Usefulness Ratings. Figure 16 shows proportions of the participants giving respective subjective usefulness rating for the two Alerts Window options. Option II received more favorable usefulness ratings than option I. The difference is statistically significant ($p = .002$).

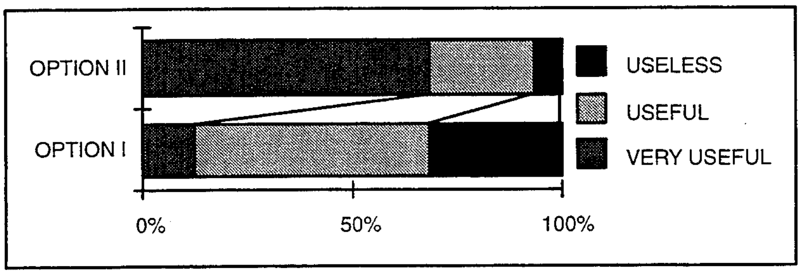

Figure 16. Proportions of subjective usefulness ratings for the two Alerts Window options.

3.2.1.2 Preferences. Seventy-five percent of the participants preferred option II, which approaches statistical significance ($p = .07$).

3.2.1.3 Interview Data. Participants provided the following recommendations during interviews.

1. Recommendations for experimental HCI. Two COs (11, 8) indicated ID and its defining factors were crucial and recommended dedicating the window to this problem. One participant (8) appreciated matching ID symbols to alert messages. Participants also suggested that display content modifications include the warning status, weapons status, and weapons posture because these affect interpretation of alerts (8). One TAO (6) suggested more concrete messages such as "turned on Cyrano 4 FC radar."

The current order for listing alerts by recency was controversial (4, 10, 16). Several participants (4, 10, 16) suggested the track and threat level as alternative ordering principles. Also, one participant (13) recommended a different organization for alert message lines: time first, track number second, and then the rest of the line. Another participant (9) suggested more highlighting of the selected track number.

Color coding alerts according to their importance was controversial. Some participants (4, 8, 14) expressed concern about a possible conflict between the color code and the air warning status. Other participants (12, 13) considered this negligible. One participant (13) said he had overlooked a white alert displayed under yellow and red alerts in the example window: the high-level alerts had masked the lower-level message. For option II, he suggested that when a more recent, lower level alert comes in, that an alert's color code stay on the screen.

Several participants felt concern about losing information and two participants (9, 10) suggested retrieval functions for the option I display. However, several participants (7, 9, 13, 14) did not like

clicking buttons. One participant (8) suggested having previously canceled alerts come back when a user deletes more recent alerts. If alerts are retrievable, the accepted number of displayed alerts was six.

For option II, a participant (16) suggested making a previously canceled track appear on the alerts list again if the user selected it again in the DSS or on the geoplot. One participant (11) suggested moving the "return to list" button, but made no suggestion on where to move the button.

2. Factors influencing preference/subjective usefulness. Participants (1, 5, 7, 9, 10, 15, 16) identified the option I display's limited capacity as an important shortcoming. Capacity restrictions also caused concern where a complete track might disappear from the option II window if a seventh track hits a tripwire (9). As one participant (15) pointed out, the superseding alerts could be an advantage if the tripwires are set appropriately, since the alerts indicate the track's importance.

Several participants (7, 9, 13, 14) indicated strong resistance to clicking any buttons. However, one participant (15) pointed out that the necessity to click on the option II history button allows the user to focus on contacts of interest. Participants (10, 15, 16) considered the history function included in option II as closer to the tactician's mental organization than just a time-sequenced list.

One CO (11) expressed concern that alerts regarding single tracks might distract him from the overall "big picture" on the geoplot. Two COs (11, 14) criticized the window's alphanumeric format and preferred a graphical attention-getting mechanism like flickering on the geoplot.

3. Use of the display. While the Alerts Window's main purpose is to direct the user's attention, one participant (10) said he would use the window to check the alerts against their priorization. Another (15) said he would use the window as an analysis tool that exceeds the attraction of attention. However, three participants (4, 6, 7) considered this analysis tool's value controversial.

One participant (12) appreciated that the Alerts Window does not overwhelm users with information. However, two COs (11, 14) said they would not use the window and would use the geoplot instead. They felt that the CO should never be distracted from the overall tactical situation. One CO (11) felt there was a dilemma: if there was no time available, he could not afford to read through the DSS. If he had time to use it, he would not need it. However, one participant (14) considered the window useful for other watch stations (14).

4. Future HCI. Several participants (1, 7, 11) raised the concern about possible alerts inflation. Current systems (including DSS) already produce up to 10 alerts per minute, so users often just turn off a system. One participant (10) appreciated that alerts do not block console access until acknowledged.

Several participants (1, 5, 10, 13, 15) considered the definition of tripwires and filters very important for the window's usefulness. Two participants (10, 13) recommended personally editable tripwires, which are ID- and platform-sensitive as well as hierarchically organized.

While two participants (2, 11) suggested integrating the alerts window with the geoplot by using pop-up messages, one participant (15) felt that researchers should not mix graphical and text items.

3.2.1.4 Questionnaire Data. Figure 17 shows the proportions of the participants' importance ratings for arguments on the Alerts Window's two options.

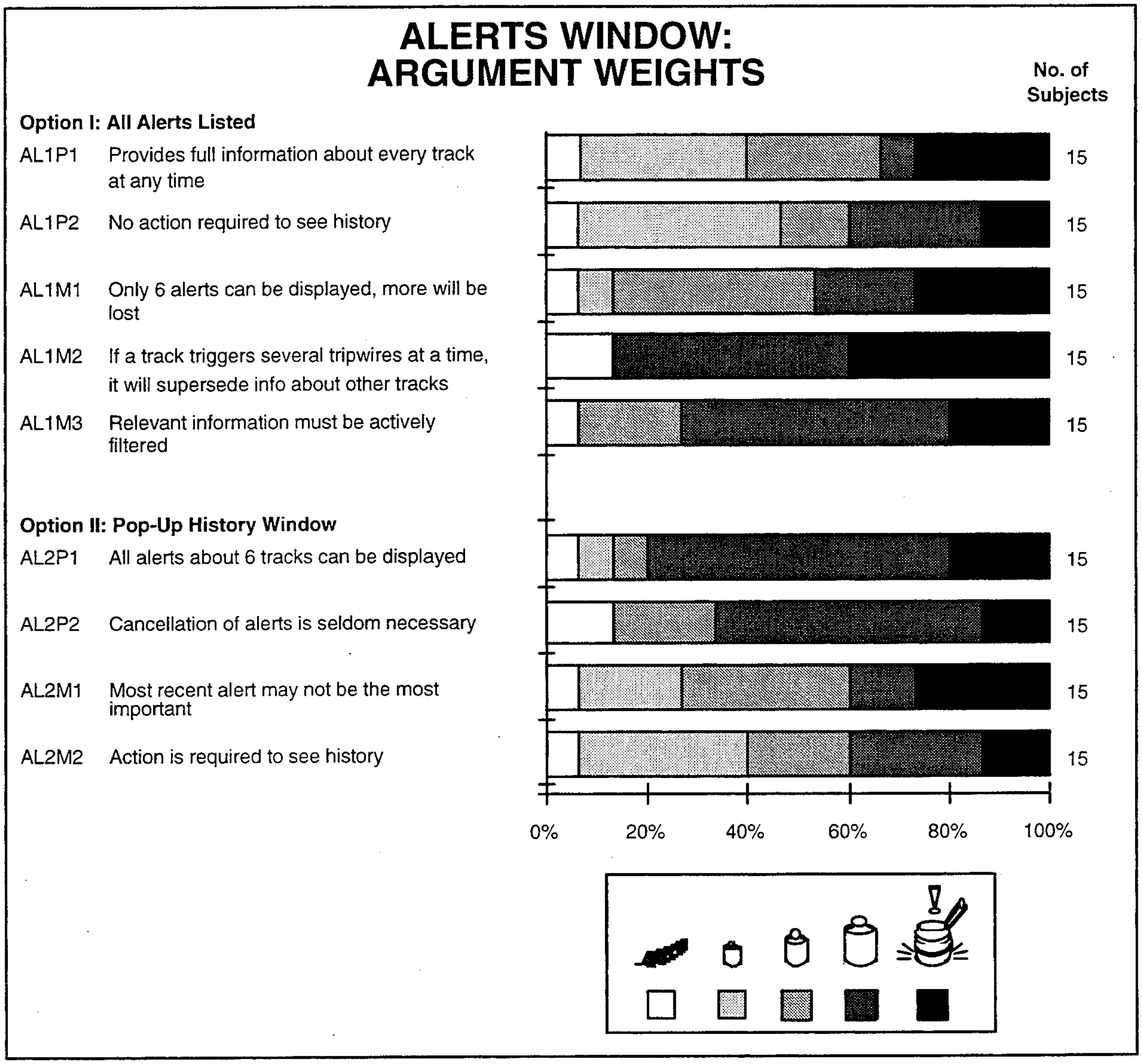

Figure 17. Proportions of participants' importance ratings for questionnaire arguments regarding the Alerts Window.

The ratings are difficult to summarize. Every argument listed received at least two "crucial" ratings, and almost all arguments were also rated as "negligible." The reason for this unexpectedly undifferentiated pattern was obviously the heterogeneity of the participants' opinions, since individual participants did differentiate between arguments. Friedman rank variance analysis confirmed this interpretation, showing significant differences between argument weights ($p = .025$ and $.123$ for option I and II, respectively), but low agreement between participants (Kendall's coefficient of concordance = 0.186 and 0.128 for option I and II, respectively).

The arguments with the most consistent and highest importance ratings were AL1M2 and AL2P1. AL1M2 stated,"If a track triggers several tripwires at a time, it will supersede information about other tracks." AL2P1 stated, "All alerts about six tracks can be displayed." Since both AL1M2 and AL2P1 consider the Alerts Window's capacity, participants considered the greater capacity of option II its predominant advantage, which is consistent with the interview data.

3.2.2 Track Profile

3.2.2.1 Usefulness Ratings. Figure 18 shows proportions of participants giving the respective subjective usefulness rating for the three displays of the Track Profile. Participants gave more favorable usefulness ratings for the altitude display than for the range and speed displays. The difference between the ratings of the altitude and speed displays is statistically significant ($p = .004$).

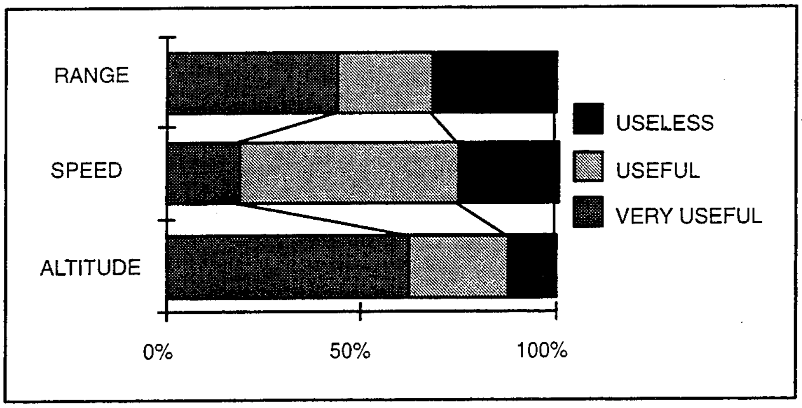

Figure 18. Proportions of subjective usefulness ratings for the three Track Profile Window displays.

3.2.2.2 Interview Data. Participants provided the following recommendations on the Track Profile Window.

1. Recommendations for experimental HCI. Several participants (2, 13, 14, 15, 16) criticized the altitude display's time scale and suggested using a range scale instead. As one participant (15) pointed out, trained users convert ranges and speeds into times, but not vice versa. Furthermore, the common range scale allowed the user to translate between the geoplot and the track profile (15). One participant (12) said that the time scale, operating with negative times, required time to become accustomed to, because zero was on the right side. An altitude over range display would further allow display of complex weapons' release envelopes (instead of release ranges alone) for track, as suggested by one participant (13) and ownship, as suggested by another (2). However, one participant (14) expressed concern that weapons release ranges could suggest a hostile intent, thus leading to premature engagement.

Two participants (13, 14) questioned the need for an explicit speed and range history graph. One participant (14) suggested adding digital display boxes in the altitude history graph, and acceleration and descent indications. He said that the window did not need more detailed speed and range history information. However, he felt that it was important to include the actual data and the trend information in the context of the altitude information. One participant (12) found a combined, fully graphical speed and altitude display "worth investigating," but another participant (14) said this would not improve usability because it would require two scales.

For the range history graph, one participant (4) recommended a polar graph of range and bearing history. This could be especially helpful in determining hostile intent of ships by maneuvering ownship. Another participant (16) said the current range over time display "takes a moment to read."

Participants (15 explicitly) generally appreciated constant scale ranges, although some participants (2, 12, 13, 14) also suggested modifiable scale ranges. One participant (8) suggested logarithmic

scales. One participant (8) said that the problem arises because there are high-altitude and high-speed weapons that require such large-scale ranges that resolution is lost. Other participants (4, 11) suggested that resolution would be lost in the low to medium speed and altitude areas, where tracks are tactically most ambiguous (4, 11). Scale range recommendations converged at 80-100 nmi for range, 50 kts speed for ships; they did not converge for aircraft speed and altitude. One participant (11) pointed out that scale ranges depend highly on the track's platform type and weapons load. However, one participant (15) said that the graph's maximum resolution was not crucial since participants imagined that users used the Track Profile to look up trends.

One participant (14) considered the time scale range too long for aircraft (14). Another (6) considered too short for ships; he suggested 30 minutes. One participant (14) suggested a time zoom function.

2. Factors influencing preference/subjective usefulness. Participants made few comments made about the window's usefulness, which seemed obvious to most participants (cf. the next section). However, one participant (3) did not "see the purpose of the window" and felt "operators may rely too much on the system." A CO (11) stated that to use the Track Profile would imply "too much focusing on one guy." User might lose the overall tactical picture if they "dive in one track's data."

3. Use of the display. Although the display principle was simple, participants outlined several different ways to use the window. One participant (13) said that the window contains well-defined missile attack profiles, but poorly defined platform attack profiles. He said "Platform behavior gives no cue whether he can shoot." Another CO (14) said he would often use the window to validate other information (e.g., from the Combat Information Center (CIC) team). A TAO (9) would use the range display first because range was his main priorization cue, while another TAO (2) felt that the tactical significance of the range display was hard to interpret. Another TAO (10) said that the Track Profile can back up functions of the option I Comparison to Norms Window. The researchers recognized the advantages of an integrated range, speed and altitude display when one participant (8) noted that the meaning of a track's descending may depend on its range. Another participant (4) suggested that descending commercial aircraft would usually slow down while an attacking craft would probably accelerate.

3. Future HCI. One participant (2) recommended an altitude versus range display with minimum ownship engagement ranges ("fire windows"). This would relate the Track Profile to the Response Manager Window. Another participant (13) suggested adding the 95% probability to hit envelopes for typical weapons and the radar horizon. This would relate the Track Profile to the Template Window. A CO (11) recommended integrating the profile with the Aegis Display System (ADS) geoplot in a 3–D display. One participant (8) recommended allowing self-defined tripwires to be set in the range display (8). Another (12) suggested user-defined scale ranges. CO (14) said that the window does not supplant the Aegis Display System's Character Readout: "Data should not be interpolated too much from a graph." Another participant (6) suggested making the Track Profile available on the AAWC console.

3.2.3 Comparison to Norms Window

3.2.3.1 Usefulness Ratings. Figure 19 shows the proportions of participants giving the respective subjective usefulness rating for the various Comparison to Norms Window options and sub-options. Participants considered Option I useful more often than any single option II display. They rated both options better than the option III display ($p = .004$ for the difference between options I and III). Option III was the only DSS window whose subjective usefulness did not differ highly significantly from 0 ($p = .125$; for all other windows and options $p \leq .016$).

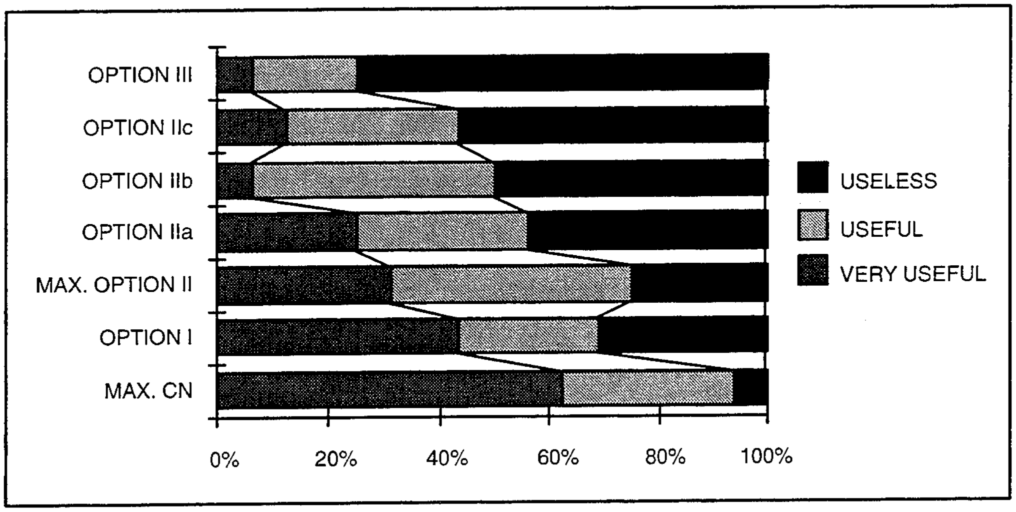

Figure 19. Proportions of subjective usefulness ratings for the Comparison to Norms Window options.

Researchers observed that participants who considered one option II (sub-option) display very useful, often rejected the other sub-options completely. Since it is not plausible to assume that the coding style of the option II matrix cells had such an impact, researchers can interpret this phenomenon as a contrast effect (see paragraph 5.2). Researchers observed a similar contrast effect for the three main options. Therefore, researchers aggregated usefulness ratings for all option II displays and all Comparison to Norms Windows, using each participant's respective maximum ratings for the aggregated (sub-) options.

These maximum ratings indicate greater acceptance for option II than for option I since there are less "useless" ratings. However, there was still more enthusiasm about option I. Participants considered this option "very useful" more often than any other option.

3.2.3.2 Preferences. Figure 20 shows the proportions of participants choosing the respective options of the Comparison to Norms Window as their strongest ("#1") preference. The p values depicted between option titles indicate the statistical significance of two-tailed sign tests on the respective sets of preference rankings.

Participants preferred option II by a slight margin over option I, and rejected option III. For option II, more than half (62%) the participant sample (figure 21; see also interview data for comments on this preference) preferred the color-coded version (a).

To test for the overall significance of these differences, Researchers conducted two separate Friedman rank variance analyses for the three Comparison to Norms window options, and option II's three sub-options. The rank sums for the main options I, II, and III were 25, 23, and 42 respectively; $p = .001$; Kendall's concordance coefficient was .484. This indicates a high statistical significance of option III's rejection, and a moderate agreement level among participants. Option II's rank sums for the three sub-options a, b, and were 22, 30, and 32 respectively; $p = .135$; Kendall's concordance coefficient was .143. Thus, there was far less agreement among participants, and the overall difference between the three sub-options' preference ranks was only marginally significant.

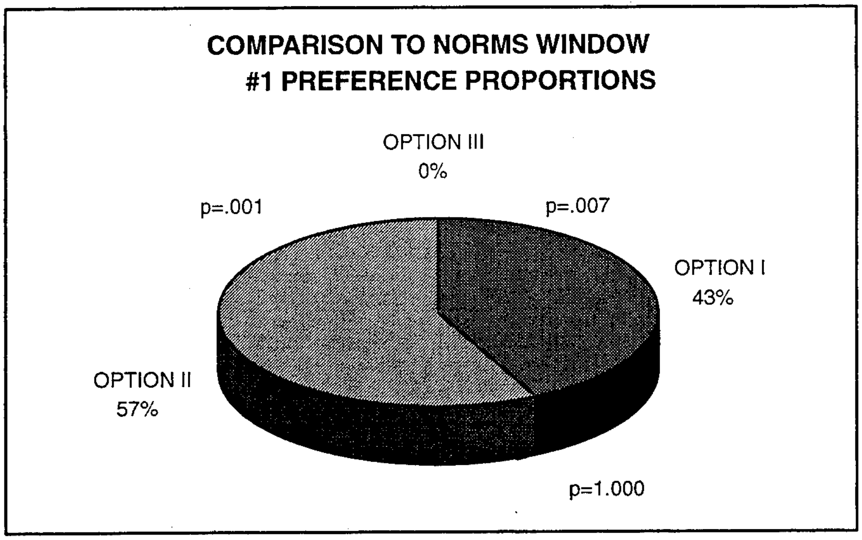

Figure 20. Proportions of participants preferring option I, II, or III, respectively, of the Comparison to Norms Window. The p values are the two-tailed probabilities of sign tests.

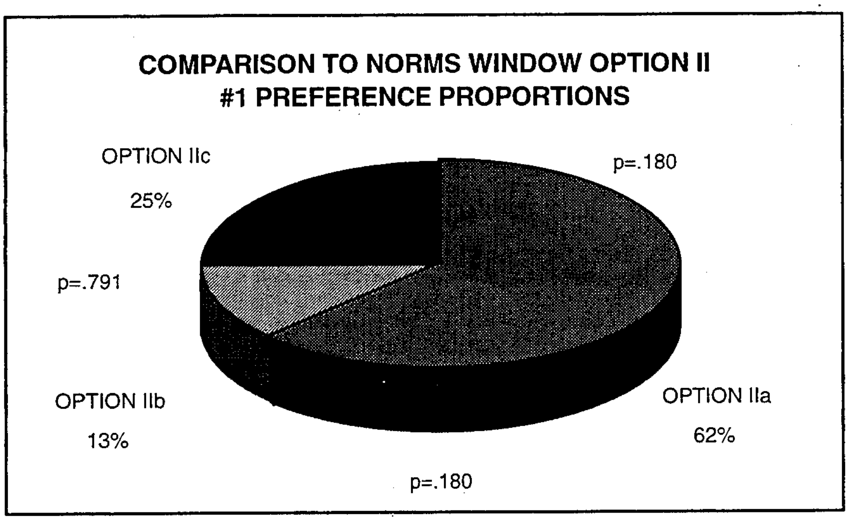

Figure 21. Proportions of participants preferring option a, b, or c respectively, of the Comparison to Norms Window option II. The p values are the two-tailed probabilities of sign tests.

3.2.3.3 Interview Data. Participants provided the following recommendations on the Comparison to Norms Window.

1. Recommendations for experimental HCI. Several participants (12 explicitly) saw the larger variable set as option II's a crucial advantage. There were even suggestions for additional variables. Several participants (2, 5, 12) recommended because high-speed aircraft can slow down while maneuvering or temporarily faking a commercial profile. One participant (11) suggested that the window include NCTR as an additional variable. Another (5) suggested renaming "Origin" to "Take off site." Actually, these are different variables. One participant (10) suggested arranging variables according to the user's personal preference (he preferred IFF at top, then altitude, speed, EW,

descent/climb rate). One participant (12) suggested adding IFF and EW data in alphanumeric format to the option I display. Another participant (7) said that a 20-minute history tail for the speed/altitude display's 20-minute history tail was sufficient.

Participants also suggested adding more platform type categories. Several participants (4, 6, 9) said the category "Fighter" was too large and suggested breaking it up into "fighter," "bomber," and "attack/tactical." One participant (8) recommended drones as an additional "Platform type" category. Another participant (11) suggested missiles.

One participant (4) suggested transposing the option II matrix because the platform type was the more pertinent information. Two participants (4,14) suggested inclining the text on the x axis to increase readability. Another (11) suggested using common abbreviations for platforms, such as VA/VF, HSL etc.

Two participants (9, 10) suggested that the option 1 display indicates that fighters are a threat by using saturated colors instead of desaturated colors. Several participants (4, 7) made the same suggestion for the option II display. One participant (7) explained that he subjectively preferred saturated colors because users might have "blurry eyes because of the bad lighting conditions" in the CIC. Several participants (15 explicitly) said they preferred color code to other coding variants in the option II display because they were more familiar with colors. One participant (8) said that option IIc's omissions looked like a screen malfunction. To avoid similar confusions, another participant (13) suggested using smaller squares instead of omissions.

2. Factors influencing preference/subjective usefulness. Several participants preferred option II because of its larger variable set (e.g.,12). One participant (12) indicated that he considered the three variables selected for the option III display the key ones, and he preferred option III over I because of the additional variables. However, option II's information density was seen as possibly disadvantageous by one participant (7), because it could overwhelm the user. Another participant (15) found option II "more usable and more dangerous." He said that pilots would not always fly within specified parameters and that the window would make more of a decision than it should. It would be dangerous to exclude platform types.

As several participants (4, 11, 15, 16) pointed out, the Comparison to Norms Window's usefulness depends strongly on the proper platform type range definition, which may be difficult to accomplish without blowing up the ranges to undiagnostic size. Some variables may not be discriminating, or very hard to determine. One participant (9) said that it was hard for the user to get intelligence data. Another (14) said time in air (additional tanks possible) was difficult, and another (11) said low-mode IFF was easy to fake. Another (6) said that an adversary's intelligence can tell a fighter pilot how to fake a non-threatening profile. As one participant (13) pointed out, "the only information differentiating between commercial and military planes is visual ID and NCTR" Being a bad guy means turning any EW off and creating confusion in order not to be shot."

The norms to compare with may be different in different areas of the world. One participant (15) preferred option I because he felt it was easier to tune for local requirements. Even though users can specify ranges as easily in the other options, he felt that the option I display gave the user a feeling for "possibilities and impossibilities," such as reconnaissance craft flying consistently lower than 10,000 feet for several days because of technical problems.

Two participants (4, 15) suggested displaying both option I and option II. Researchers could use option I as a standard, and they could use option II if participants requested additional information.

One participant (15) pointed out that several variables in option II are cumulative over time. Thus, option II might be more useful (than option I) later in the process. For new tracks it would only display speed and altitude, which option I did better. While one participant (8) explicitly stated that he considered all options to be useful, another (13) said none would be necessary, and he would use them "as an afterthought only."

While one participant (9) preferred option I "because it takes less time to look at," another (11) preferred option II for its "computational speed." One participant (12) explained that the missing relations of the track's data to norm boundaries (such as below, beyond, close to...) are not a strong disadvantage of option II as long as they can be looked up by clicking on the respective cells.

Most participants disliked option III. One participant (15) stated explicitly that it was "extremely hard to use and takes forever." He also said that looking at different variables involves memory load, which was particularly bad under high stress conditions.

3. Use of the display. Several participants (6, 10, 12) felt that the option I display is partially redundant with the Track Profile since both display altitude and speed history. Some participants (6, 10) preferred option II because they could derive the option I information from the track profile, however, not vice versa. One participant (6) even felt "option I's value comes from the parts that are redundant with the Track Profile." One participant (10) said he preferred option I if used alone, and option II if the user could access the Track Profile. Another participant (12) felt the redundancy with the Track Profile was not necessarily a disadvantage because "the user does not lock on one display."

Option II's use is not obvious, but two COs (8, 12) said this would not be a problem because users would have months of training and experience with the system. One CO (4) said that the difficulty is that "a column that is half yellow is as diagnostic as a full green column," because both columns still indicate possible platform types. Two participants (15,16) felt that the display might suggest a "best" classification solution (a column with many "fitting" cells), which is not better than an alternative solution supported only by "uncertain" cells. Another participant said that the display did not readily inform the user on why it excluded a platform type. One CO (11) said that he had previously tried to use a similar matrix for the ID problem, but it turned out in training sessions that "good guys could be shot" if the user followed the matrix's suggestion blindly. However, another CO (8) pointed out that the option II matrix was "what you do anyway." Certain schools did teach this way of processing information.

The option II display's multiple data integration gives the user additional possibilities. One participant (6) said that the EW information would be also available if the EW operator is busy. Another participant (10) said that the overall data set would allow the user to "get a better feeling about the team's decisions," i.e., to monitor the CIC team's effectiveness.

4. Future HCI One CO (14) felt that the option I display was not useful for the CO, but was useful for ID operators. It might be dangerous for the CO because it should not be the only tool used. The option II display would possibly be a separate watch station operated by an enlisted man who would requiring extensive training.

One participant (15) suggested tuning the Comparison to Norms Window (option II) to the ID problem, which he considered to be much more important than the platform type. He suggested using the variables track profile, country of origin, EW, IFF, ID maneuver, return to forces (RTF) profile, and tactical air corridor use.

3.2.3.4 Questionnaire Data. Figure 22 shows proportions of participants' importance ratings for arguments concerning the three options of the Comparison to Norms Window.

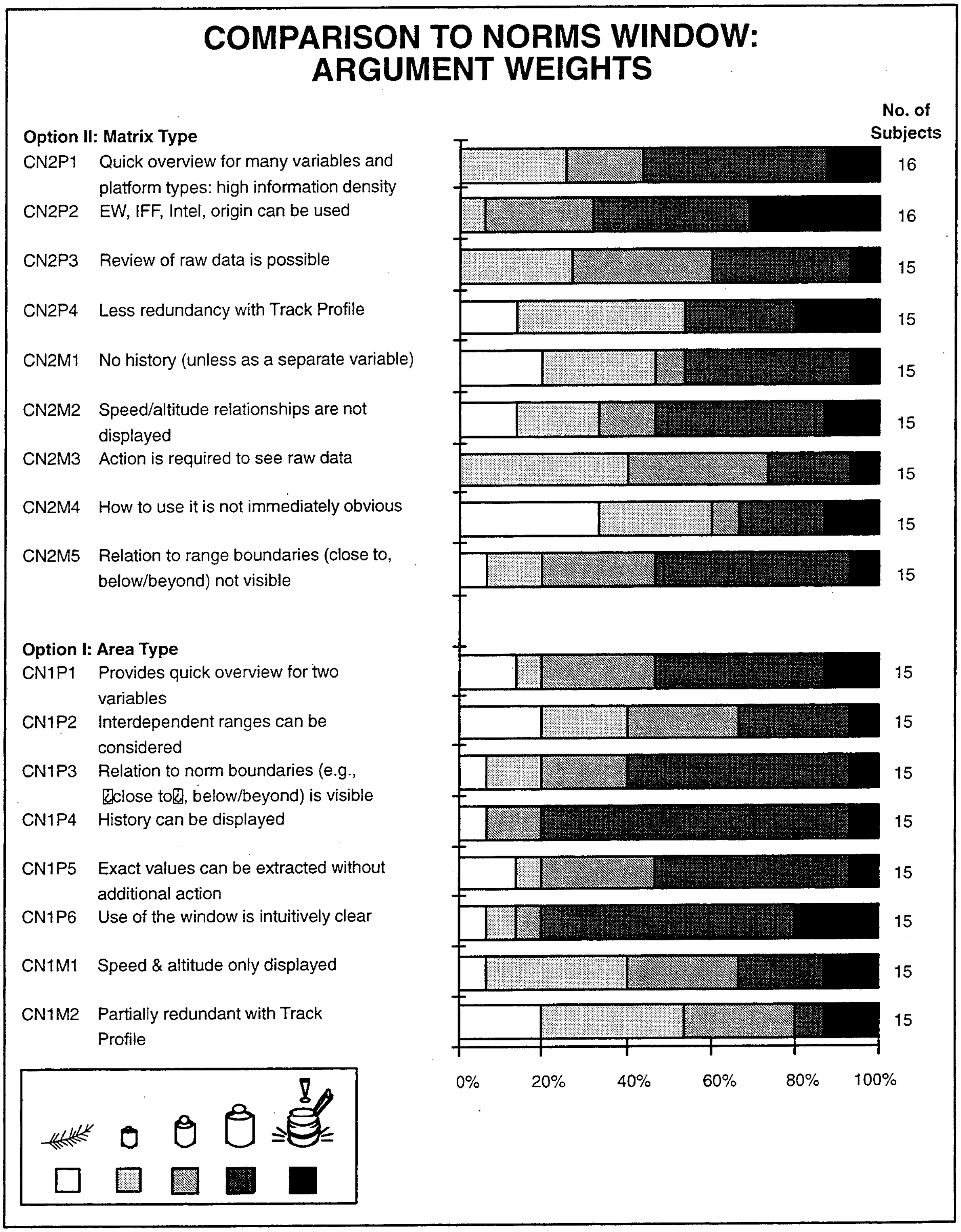

Figure 22. Proportions of participants' importance ratings for Comparison to Norms Window questionnaire arguments.

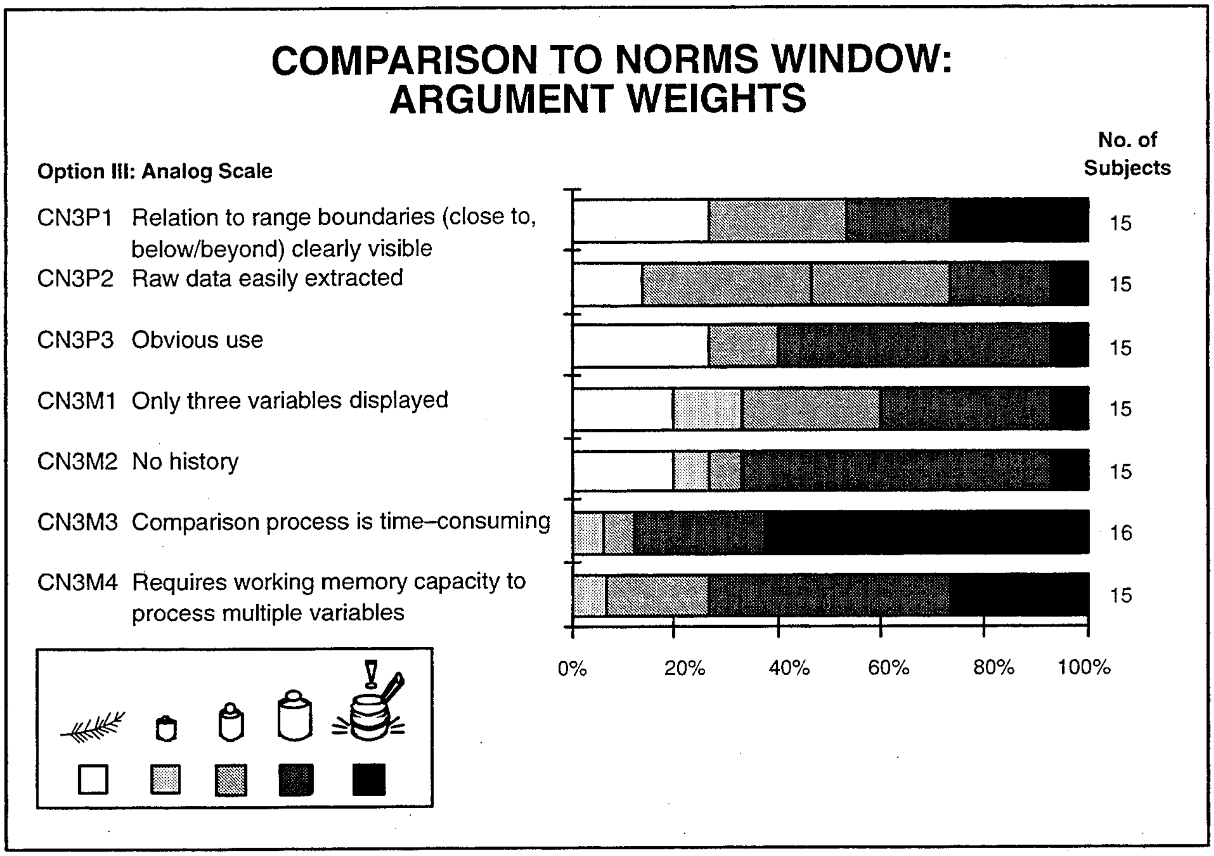

Figure 22. Proportions of participants' importance ratings for questionnaire arguments regarding the Comparison to Norms Window. (continued)

As in the Alerts Window, argument importance ratings were very heterogeneous, with all arguments but two receiving "crucial" ratings. Again, other participants also rated most arguments as "negligible." Accordingly, a Friedman rank variance analysis showed significant differences between argument weights (p = .016, .107, and .000 for option I, II, and III, respectively), but few agreements among participants (Kendall's coefficient of concordance = 0.165, 0.107, and 0.282 for option I, II, and III, respectively).

Option I's highest-rated advantages were its intuitive use (CN1P6) and the history display (CN1P4). Note that the latter is not a genuine feature of the window. For option II, participants consistently considered the quick usability of high information density (CN2P1 important. However, participants rated the access to additional data (i.e., EW, Intelligence, IFF, and origin; CN2P2) important more often. Thus, researchers considered the amount of information presented at least as beneficial as the display format that integrated this information. They consistently considered option III's demands on user's time (CN3M3) and memory capacity (CN3M4) to be important shortcomings. The high importance ratings of these two items account for the participants' moderately higher concordance on this issue.

3.2.4 Template Window

3.2.4.1 Usefulness Ratings. While 12.5% of the participants considered the Template Window very useful, 62.5% found it useful, and 25% useless.

3.2.4.2 Interview Data. Since there were no different options to discuss, the "early" participants did not make many comments on the Template Window. Thus, researchers asked participants for comments on this window later in the interview series.

1. Recommendations for experimental HCI. After one participant (14) suggested a template based on a range scale instead of a time scale, researchers asked the following participants directly about this issue. All agreed that the range scale would be better than the time scale. One participant (14) said that the main reasons were consistency with the geoplot and nominal attack profiles, and that "tactics are based on range, not time." These arguments are identical with the previously stated advantages of a range-based Track Profile Window. One participant (15) even suggested a graphical altitude over range display similar to a range-based Track Profile. This display would include "activity areas," since certain activities in an attack sequence depend not only on range, but also on altitude. Users would see activities like "descending" and "turning inbound" immediately. The display would not label these areas with text. Another participant (16) found this display concept "much easier to absorb" than the alphanumeric display, and "a lot of information for a guy just walking in."

Two participants (10,14) recommended that the "attack" hypothesis be renamed to "something less aggressive." One participant (10) recommended this because "inexperienced TAO trainees tend to be aggressive beyond ROE" (10). Another participant (14) disliked the hypothesis plausibility coding by the position of the respective button and suggested highlighting or color coding. Another participant (13) recommended interpolation of the template's time bars' movement between data updates when the update rate is one or less per minute, since the window could otherwise be very misleading.

2. Factors influencing preference/subjective usefulness. Four participants (3, 4, 11, 14) indicated explicitly that they disliked the window. As one of these participants (14) pointed out, the template to display would depend on the track's ID. He would have no confidence in the system because ID was hard to determine. He also explained that he lost confidence in the window's usefulness because of the example window's inconsistent and contradictory content. Furthermore, he did not see the need for the template because the Track Priority List displays the three most likely hypotheses.

Another CO (11) said that he "would never go this deep for one track" because losing the overall tactical picture would concern him. He said he would try to obtain a visual ID anyway if the track's ID was unclear.

3. Use of the display. One participant (12) appreciated that the Template Window provided a default hypothesis that is also the most likely one. However, given the overall picture of participants' comments, the main question seemed to be whether the track was attacking or not (14, 16). Two participants (12, 10) pointed out that if it was attacking, the expected behavior template depended strongly on the weapons carried. Only visual ID and intelligence information could determine weapons carried. Another participant (14) suggested that for tracks not attacking, users had only minimal or no interest in intent.

Several participants (10, 13, 16) said that the difference between attack and harassment was very fuzzy and hard to determine. Consequently, participants' suggestions differed on how to differentiate. One participant (10) felt the only valid information on this issue would be weapons carried by the track. A second participant (11) would only accept visual ID, NCTR and a lack of mode 4 IFF, if the same aircraft or section detected them, to positively determine hostile ID. Another participant (13) noted the differentiating logic would possibly be counter-intuitive. A track on a harassment mission would try to look quite dangerous with a close closest point of approach (CPA). An attacking track in cold war would try to hide its intentions as long as possible by flying a non-threatening profile, faking an air distress or even using a reconnaissance craft. Also, the first pass might be for

harassment, but the second might be an attack. Thus, intercepted communication, intelligence information, and weapons carried would be the best information differentiating between attack and harass. Since the Template Window includes none of these data, several participants (3, 4, 11 explicitly) questioned its usefulness.

Two participants (14, 15) indicated that they would use the window as a checklist, looking for the white triangles (indicating a match between the template and actual behavior). One CO (16) implicitly suggested using a modified, graphical altitude/range Template Window as a tool for briefing a CO called into the CIC in a critical situation.

3.2.5 SABER Window

3.2.5.1 Usefulness Ratings. Figure 23 shows proportions of participants giving the respective subjective usefulness rating for the two SABER Window options. Statistically, the difference is only marginally significant (p = .125).

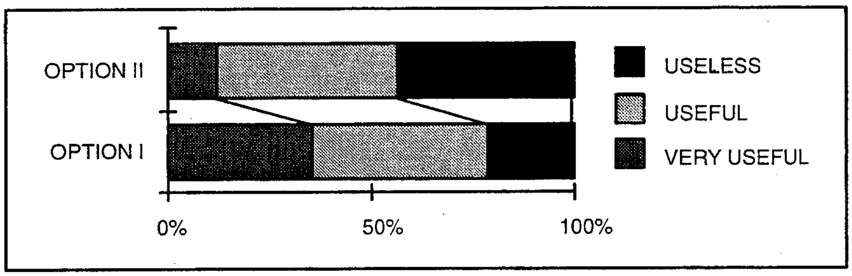

Figure 23. Proportions of subjective usefulness ratings for the two SABER Window options.

3.2.5.2 Preferences. Eighty-two percent of the participants preferred option I, which the researchers cannot consider statistically significant (p = .15) due to the small number of participants (N = 11). The SABER Window (option I) was not available for the first four participants.

3.2.5.3 Interview Data. Participants provided the following suggestions on the SABER Window.

1. Recommendations for experimental HCI. Two participants (9, 15) suggested either dimming evidence items that are common across hypothesis categories or listing them separately from the single evidence lists. One participant (5) suggested a toggle switch to filter/unfilter the evidence lists. Another participant (7) suggested that this switch filter depending on the user's familiarity with the system, or on the situation, because, for instance, attack acceleration looked different from air distress acceleration. Another participant (15) suggested users should form lists that would force them to think intensely about hypotheses.

Three participants (10, 11, 13) made suggestions about the evidence list content. According to these participants, the lists should include information about weapons load, ID, and IFF. The track's flying triangles could identify air distress situations.

2. Factors influencing preference/subjective usefulness Whether or not to filter items that are common across hypothesis categories from the evidence lists was controversial. Two participants (7, 16) explicitly preferred the filtered list. They felt this because common items "do not contribute to the picture" and "looking at additional information does not help verify the system's suggestions." Other participants (1, 9, 10, 15) felt that the system should not withhold information. One participant said that the user should know the information filtered out to create the filtered list.

While one participant (5) stated, "The more you look at it the better you like it," other participants (11 explicitly) stressed intent's relatively small importance to ID. They felt intent reasoning drew attention away from ID classification, which they considered much more important than intent, since ROE are purely based on ID. Finally, users could derive ID from intent. As another participant (15) noted, "to focus on intent presumes known ID." Since the ID problem was so difficult, two participants (4, 11) said that they would never trust a machine.

Two participants' (11, 13) comments reflected the fuzziness and difficulty of the attack/harassment differentiation already mentioned in the Template Window section. One participant (13) further specified: "In cold war an attack would probably look very un-threatening. One guy attacking a battle group would try to hide as long as possible. On the other hand, a terrorist could behave in any way. Behavior would be very culture-specific and differ from the Gulf area to North Korea. Military attacks would differ from terrorist attacks."

One participant (11) noted that users needed time to use the window, and if they had time, they would not need it. He also felt that the evidence list boxes were too small, but if the SABER Window would hold all relevant variables, it would clutter the screen.

3. <u>Use of the display.</u> Several participants imagined using the window differently than its intended use. One participant (6) stated he would use the window to look up all information the system holds about a track. One participant (5) suggested using the window to brief the CO about a given situation. Another participant (13) suggested using the window to get background information about the tactical situation during a watch's first 45 minutes. A skeptical participant (14) felt that the window was useful only for "academic reconstruction exercises" or to verify the algorithm determining the hypothesis. This participant also stated explicitly that he would not test assumptions on alternative hypotheses, since CO and TAO "decide upon accepting the given ID, but they do not work on the ID problem."

4. <u>Future HCI.</u> One participant (2) said that the window's usefulness might mask its position in the screen's lower right corner. He recommended that a pop-up window become active on a hooked track. Another participant (15) said that he had previously tried to use tactical tripwires and insisted that they be editable by an ownship operator. Different track categories might have a common action set, and the problem would be to determine relevant information for selecting an action.

3.2.5.4 Questionnaire Data. Figure 24 shows proportions of the participants' importance ratings for arguments concerning the two SABER Window options. When comparing option I's three arguments, researchers found no significant differences in subjective importance ($p = .67$ in a Friedman analysis of rank variance). However, on the individual participant level, this result did not hold. Individual participants did not agree on the importance ratings (Kendall's coefficient of concordance = .031).

Option II's most important arguments (significantly; Friedman $p = .008$) were related to a secondary function of the SABER window, i.e., the availability of information to verify the system's operation. However, participants did not feel that it was necessary to verify the system's performance, but highly desirable *to be able* to do so. The two arguments rated most important were disadvantages of the filtered evidence list. Both arguments relate to the user's trust in the system. SA2M1 concerns a user's inability to verify the system's database, and SA2M2 concerns a possible pitfall in the filtering algorithm. Participants moderately agreed on these arguments (Kendall's coefficient of concordance = 0.304). However, note that participants considered the corresponding advantage of option I, that users can look up all evidence processed by the system (SA1P2), less important. Thus, they considered the possible loss of verifying information more important than its availability.

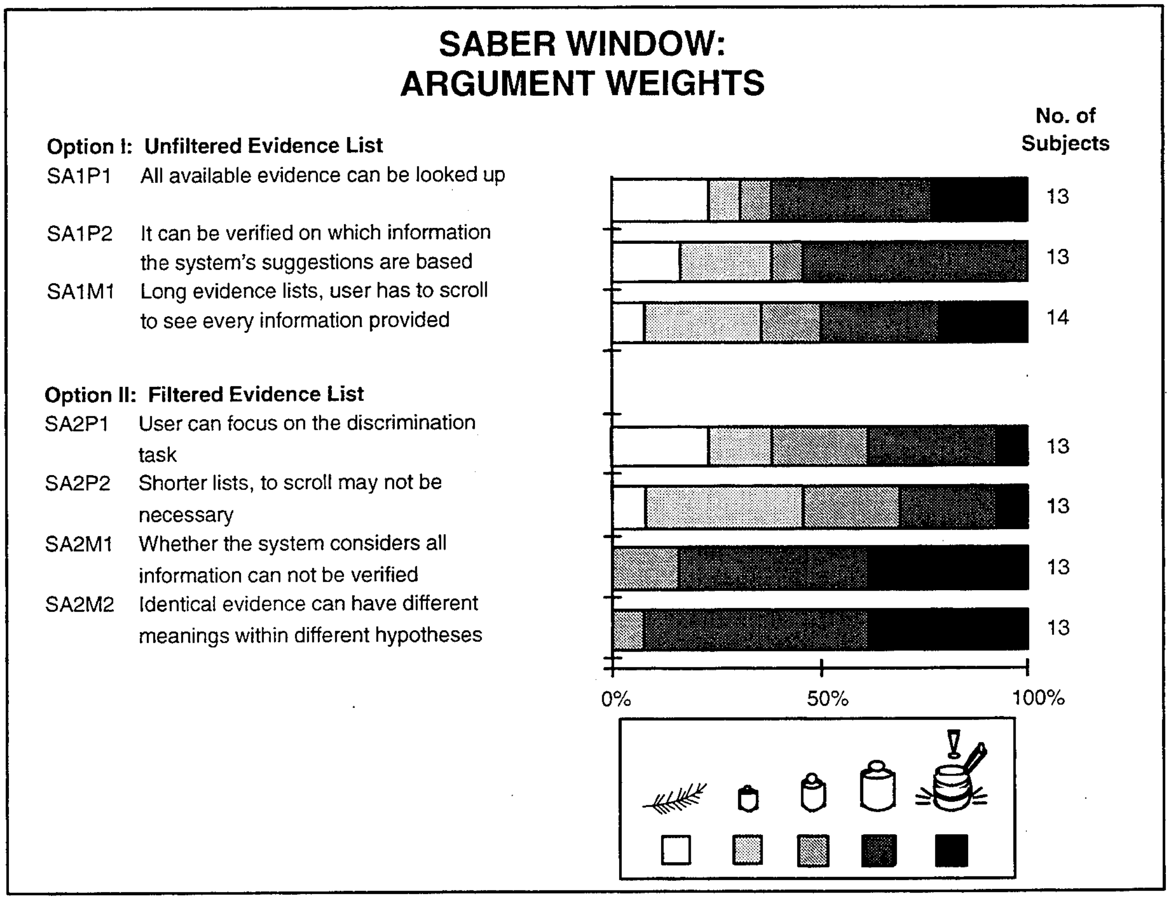

Figure 24. Proportions of participants' importance ratings for questionnaire arguments regarding the SABER Window.

3.2.6 Response Manager Window

3.2.6.1 Usefulness Ratings. Usefulness ratings for the Response Manager Window cannot be discussed independently from the ROE support format chosen. Therefore, this report discusses them in the ROE display section (paragraph 3.2.7.1).

3.2.6.2 Preferences. Eighty-seven percent of the participants preferred option II, which is highly significant statistically (p = .004).

3.2.6.3 Interview Data. The participants provided the following recommendations on the Response Manager Window.

1. Recommendations for experimental HCI. While there was a strong preference for a one-dimensional arrangement of the response plan, there was some controversy about whether the system should use a time scale or a range scale. One participant (13) tied ROE and battle orders to ranges, not to time. Using a time scale would unnecessarily go "into micro management." One participant (14) especially liked option I in that it displayed the track's actual range. However, one participant (16) accepted the time scale. One participant (14) suggested basing the Response Manager Window on the same display format and symbology as the Template Window, but in separate windows.

Several participants (2, 4, 5, 6, 7, 9, 12) appreciated option II's feedback function. Some participants (2, 4, 5, 6, 7, 9, 11, 12) also emphasized the need for automated feedback about actions taken

and missed. One participant (6) said that users expect manual feedback to increase their workload. Another (7) said that it could possibly increase errors (7). One participant (15) pointed out the amount of feedback messages would be difficult to handle: six actions within 3 minutes are two actions per minute. If there were only three tracks, the screen would have to provide feedback every 10 seconds. Therefore, he doubted whether feedback would ever be accurate. One participant (2) suggested that feedback should indicate the time when an action has been taken. While one participant (4) suggested that option I should provide feedback using color code (4), another (7) suggested providing feedback by collapsing the respective action angles, which would also increase the display's readability.

Participants made several comments on the actions to be displayed. Researchers forwarded these comments to the subject matter experts who ultimately define a window's content. Researchers grouped the comments into three classes, concerning either within-ship operations, or legally required operations that "leave the ship" (i.e., communications, alerts, and EW measures), or reports to senior officers. Researchers saw the priority of these reports in various ways.

Single recommendations concerned the display format. One participant recommended including a finer time scale for fast attack fighters. Another participant (6) suggested color-coding ROE-based actions because the law requires it. A participant (9) suggested increasing the option I display's readability by reasonably scrolling the window along the range scale while keeping the range pointer in a constant position.

One participant (13) said that three strategies would not be necessary since the Response Manager Window should display only legally required responses. Another participant (11) recommended searching for a better word for "deconflict" because it could also mean to deconflict friendly forces' weapons and planes.

2. <u>Factors influencing preference/subjective usefulness</u>. Although two participants (12, 14) appreciated the combination of time and range scales, most participants (11, 12, 15, 16 explicitly) considered option II's time line sufficient and preferred it for its better readability. Researchers assumed that most participants would prefer a one-dimensional scale for response arrangement. One CO (11) said that the Response Manager Window led the user away from the air picture, making him focus too much on a single track.

3. <u>Use of the display.</u> Two participants (7, 10) said that some ships already use lists for pre-planned actions similar to the Response Manager Window. Two participants (7, 10) said the ships use the pre-planned actions list much like a checklist for engagement procedures. One participant (10) indicated he would also use the Response Manager Window as a memory backup. Several participants (7, 10 explicitly) said that they could use the window to brief a CO just called into the CIC for a critical situation. This was one of the reasons why participants widely appreciated option II's feedback function. One participant (7) suggested time tagging actions and printing them out as an action report.

4. <u>Future HCI.</u> Seven participants (2, 4, 5, 6, 7, 9, 11) suggested automating the feedback function by having messages sent from the effector side, i.e., other CIC consoles. One participant (11) suggested integrating the Response Manager Window into a 3-D tactical overview, but wondered where to put it.

3.2.6.4 Questionnaire Data. Figure 25 shows the proportions of participants giving the respective importance rating for arguments about the Response Manager Window used in the questionnaire. As

in most other windows, almost every item received "crucial" ratings. However, differences in subjective importance of the individual arguments were statistically significant (Friedman $p = .056$ and .003 for option I and II, respectively). Again, participants' agreement on argument importance was low (Kendall's coefficient of concordance = .146 and .232 for option I and II, respectively).

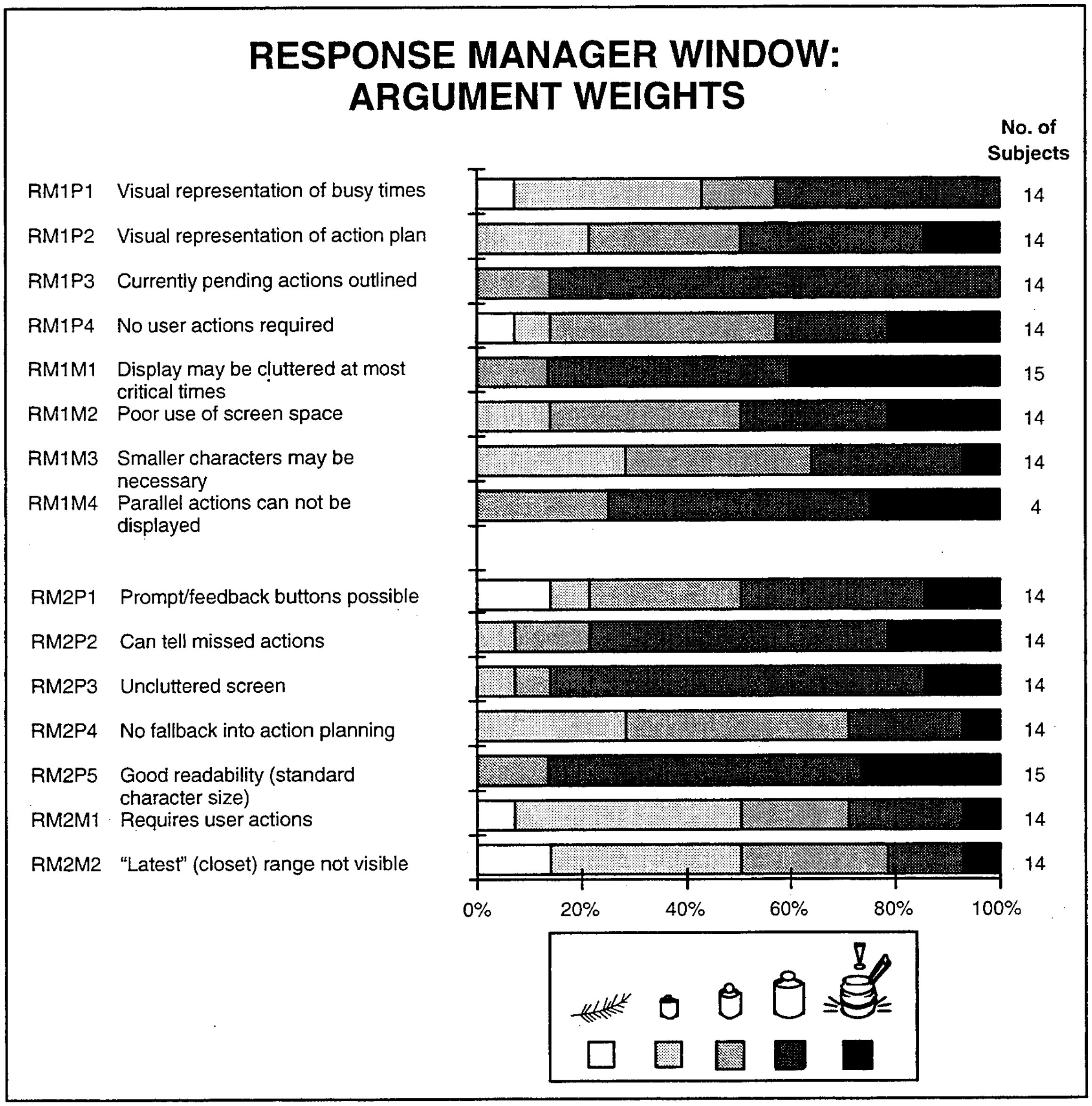

Figure 25. Proportions of participants' importance ratings for questionnaire arguments regarding the Response Manager Window.

The most important arguments concerned the display's legibility (RM1M1, RM2P5, RM2P3), which is consistent with interview comments. A further subjectively important advantage of option II was the availability of feedback (RM2P2). Statistical analyses did not include RM1M4, "Parallel actions cannot be displayed [in the option I display]," because of its late introduction in the questionnaire. However, the four subjects asked considered it consistently quite important. The argument reflects that actions planned at the same range would supersede each other in the option I display.

3.2.7 ROE Display

3.2.7.1 Usefulness Ratings. Figure 26 shows proportions of participants giving the respective subjective usefulness rating for the Response Manager Window options, combined with the ROE support.

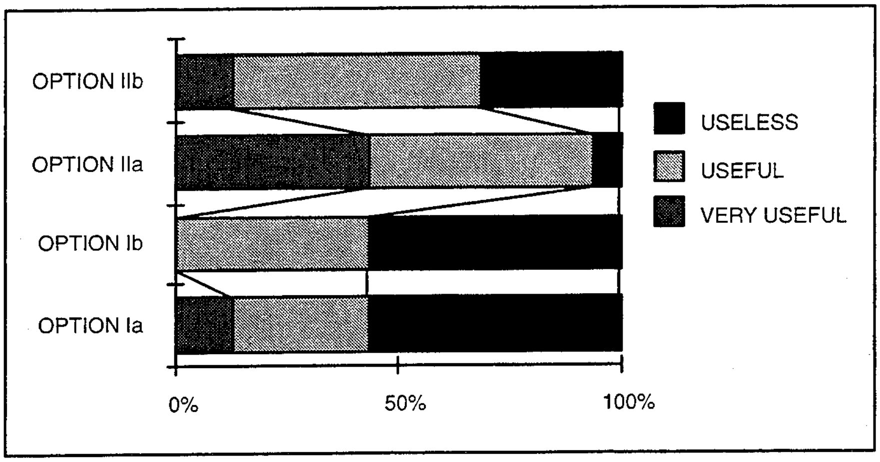

Figure 26. Proportions of subjective usefulness ratings for the Response Manager Window options, combined with ROE support option.

Participants clearly considered option IIa, the time line display with the ROE box the most useful. Generally, participants considered this display useful significantly more often than both option I displays ($p \leq .012$). A contrast effect similar to the one discussed for the Comparison to Norms Window may overshadow the participants' opinion. For the ROE support, the ROE box's greater subjective usefulness, compared to integrated ROE, was not as clear. Only within the option II Response Manager Window, did participants express a marginally significant difference between option a and b ($p = .146$).

3.2.7.2 Preferences. Eighty percent of the participants preferred the ROE box separate from the action plan (option a), which is statistically significant ($p = .013$).

3.2.7.3 Interview Data. Participants provided the following suggestions about the Response Manager Window.

1. <u>Recommendations for experimental HCI.</u> Several subjects (1, 5, 6, 7) disliked the checkmarks in the ROE table (cf. "use of the display"), feeling they could lead to premature engagement.

2. <u>Factors influencing preference/subjective usefulness.</u> Several subjects (1, 4, 6, 8) perceived ROE guidance as separate from the concrete action plan, making a comprehensive integration impossible. However, some participants (10, 11) considered it possible that a user might ignore a ROE box separated from the action plan. ROE integration requires proper feedback about actions completed and missed. One participant (6) suggested, "This will require interaction which may distract from a fast moving tactical situation." Another participant (11) questioned the need for any ROE display. He said that in DSS situations the weapons status is never higher than "tight." In the gulf war it was never higher than [air warning] yellow, safe. Users always have to tell the CO always to engage, unless in a clear self-defense situation.

3. <u>Use of the display.</u> One participant (4) perceived ROE guidance as separate from single items on the action plan. Another participant (6), representative of most participants, considered the ROE

display to be a "nice tool for reminders." He felt applying ROE was a dynamic event: "You may not engage even if the ROE rules have been met."

Several participants (1) explicitly, see also questionnaire data) pointed out that trained TAOs: "Don't let ROE become a checklist!" A CO (8) said that rules of engagement are "incompatible with digital thinking." However, another participant (7) remarked that rules of engagement are kind of a checklist in situations close to an engagement. The rules do not give permission to shoot, but they must be met before shooting. Furthermore, as another participant (10) pointed out, checklists almost thoroughly guide anything leading to ROE-based decisions, making the process finally a checklist decision. One participant's (13) statement summarized the discussion: "ROE is a checklist. The CO controls the TAO to prevent premature actions. When the user meets the rules of engagement, the track's analysis done."

4. Future HCI. One participant (6) indicated that he would welcome any effective ROE support, and pointed out several reasons why this is also very hard to accomplish. Note that the researchers set up experimental ROE for experimental purposes. The experimental ROE was much more concrete and less complex than "real world" ROE. The following discussion does not apply to the experimental DSS.

Several participants (4, 6, 11, 16) indicated that they did not think a small table can summarize real-world ROE. One participant (8) said that the ROE were not written in concrete numbers, but gave general advice for several possible courses of action. Therefore, they were probably impossible to integrate in the action plan. However, this was not even desired by most participants (8, 4, 15), who felt ROE offered general guidance separate from the action plan (although related). They felt that it was necessary that the CO actively integrate ROE into the action plan. Another participant (14) said that a ROE box "displays facts, while the Response Manager [action plan] displays tasks."

Two participants (7, 12) suggested displaying ROE in a context-sensitive way, i.e., to display engagement-related ROE only if engagement becomes an option. They felt that a small display box was sufficient for a case-sensitive ROE display since ROE are often platform-specific. Another participant (13) remarked that usually there was no sequence in ROE and thus a small "remember" box would suffice.

One participant (13) said that multiple, separate sets of ROE, e.g., for different countries, are not uncommon. Another participant (5) said that the user needed control over the ROE table contents.

3.2.7.4 Questionnaire Data. Figure 27 shows proportions of participants giving the respective importance rating for ROE support function arguments used in the questionnaire. Differences in arguments' subjective importance were not significant (Friedman $p = .409$ and .225 for option a and b, respectively), and agreement among participants was low (Kendall coefficient of concordance = .124 and .177 for option a and b, respectively). However, the three arguments rated most important (ROE1P1, ROE1M4, and ROE2M3) were consistent with the interview data.

3.2.8 Track Priority List

3.2.8.1 Usefulness Ratings. Figure 28 shows the proportions of participants giving the two Track Priority List options' respective subjective usefulness rating. Option II received more favorable usefulness ratings than for option I. However, the difference was not significant statistically ($p = .388$).

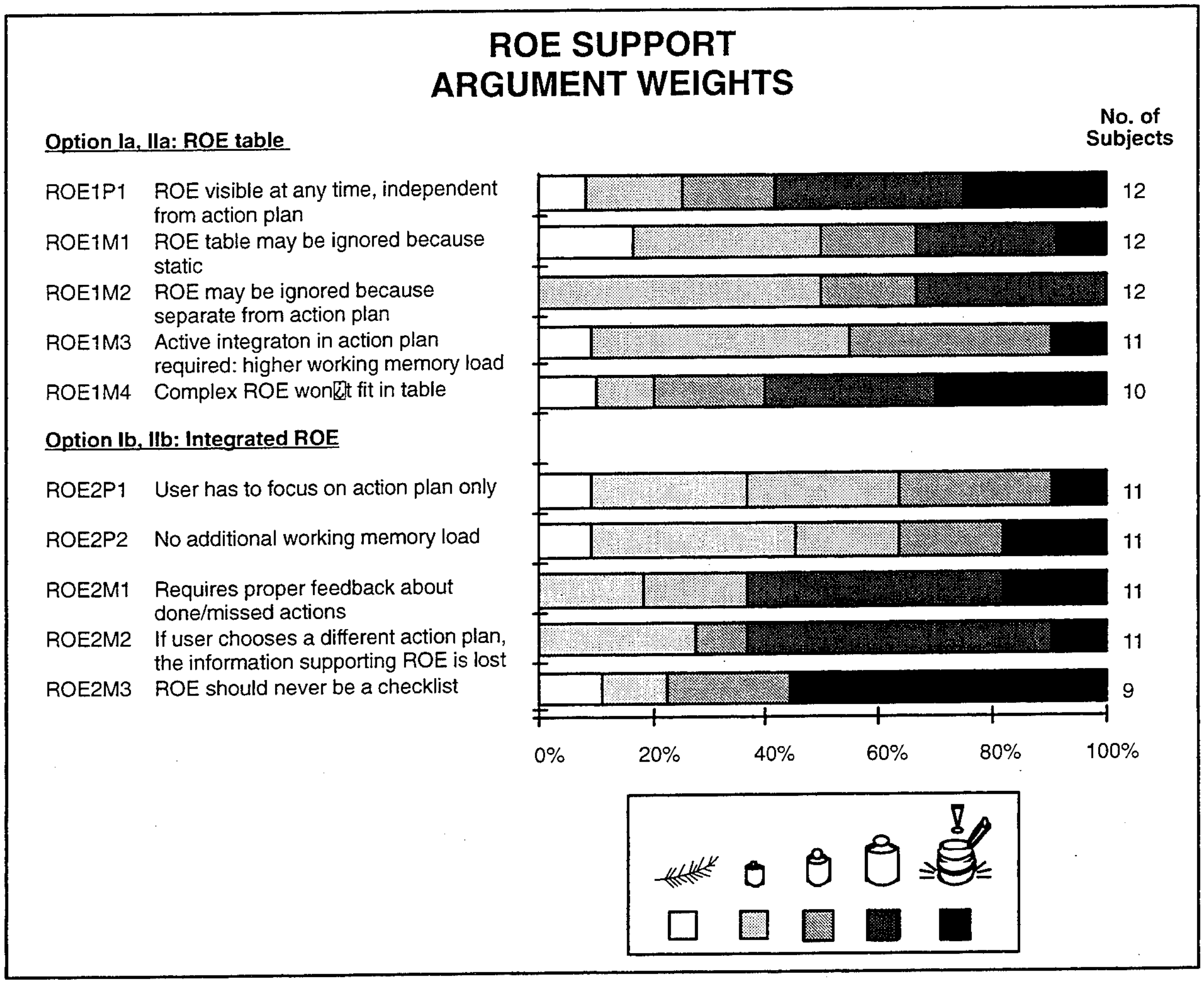

Figure 27. Proportions of participants' importance ratings for questionnaire arguments regarding ROE support.

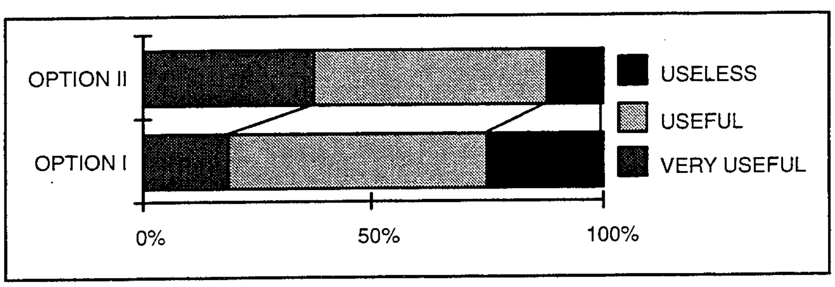

Figure 28. Proportions of subjective usefulness ratings for the two Track Priority List options.

3.2.8.2 Preferences. Fifty-three percent of the participants preferred option II (bearing display), which is only a slight margin with no statistical significance.

3.2.8.3 Interview Data. Participants provided the following suggestions on the Track Priority List.

1. Recommendations for experimental HCI. Participants made several suggestions concerning the track priority definition. One participant (11) suggested the current Aegis logic, i.e., to order tracks

on the list by the last opportunity to engage. Two participants (13, 14) recommended going beyond this and ordering the list by a combination of ID (hierarchy: hostile > assumed enemy > unevaluated > assumed friendly), time until the track reaches its weapons release range, and last opportunity to engage. One of these participants (14) also suggested the track's profile as a factor, while the other participant (13) pointed out the priority assessment also depended on the region and current political situation. Another participant (16) considered the priority model unimportant since the window's capacity was large enough to hold all priority tracks.

One participant (15) suggested a completely different concept for the Track Priority List. He considered automatic track priorization unrealistic because of the usually unknown, widely varying weapons release ranges of tracks. He suggested using tripwires similar to those in the Alerts Window that should have a "smart hierarchy," and selecting tracks manually for the Priority List. He also suggested manually entering possible weapons and assumed intent, and following the actions recommended on this basis by a Response Manager Window.

Four participants (4, 9, 10, 14) recommended adding the range to the bearing information, since they "belong together." Another participant (14) suggested adding the tag. One participant (2) said the tag should be entered automatically, while another (9) said the tag should include more characters. Others (6, 12) recommended a more intense use of color for items like "engage" because they felt color was "the biggest help for snapshot glances at the screen."

Two participants (4, 11) suggested displaying alerts messages within the Track Priority List. One of these participants (11) suggested replacing the intent buttons (11). Another participant (14) felt the Track Priority List should be dedicated to weapons and engagement only and suggested taking out "administrative actions like reporting."

2. <u>Factors influencing preference/subjective usefulness.</u> Most participants preferred having the bearing displayed on the Track Priority List instead of the tag. As one participant (13) explained, the bearing gives an identification cue before the track is tagged, which is important for fast tracks. Another participant (10) said he used the track number when he tagged a track so the tag display would be redundant. So, even though track numbers can be called up on the Aegis console (6, 10,16), participants appreciated the additional orienting support.

Two participants (7, 15) commented on their preference for the tag display. One (7) pointed out that the bearing information is available elsewhere. The other participant (15) preferred the tag because it indicated priority tracks independently from the system's priorization. Note this participant was very skeptical about the usefulness of an automated Track Priority List (see the next paragraph).

There were three general types of criticism. One participant (10) said that in the past, priority lists were usually too slow and never up to date, which spoiled their functionality. Another participant (11) said that users should focus on the geoplot because "people that look at alphanumerics are dangerous to their own guys." The third criticism was from a participant (15) who said that he considered automated priorization impossible to accomplish in a meaningful way as long as a track's weapons release ranges are unknown. Automatically derived assumptions were usually too negative to be useful

3. <u>Use of the display.</u> One participant (8) pointed out that the bearing information could be particularly useful for the communication between CO and TAO. The TAO knew where a track with a given number is, but the CO maybe did not. Similarly, one TAO (10) said he would use the Track Priority List and the Response Manager Window to brief the CO about the tactical situation.

4. Future HCI. One participant (2) noted that problems might arise if the track number changed, which was not uncommon if tracks are temporarily lost. The same participant said he would appreciate an automatic tagging function.

3.2.8.4 Questionnaire Data. Figure 29 shows proportions of participants giving the respective importance rating for the Track Priority List's arguments used in the questionnaire. Differences in the listed arguments' subjective importance were not significant (Friedman p = .489 and .205 for option I and II, respectively), and agreement among participants was low (Kendall coefficient of concordance = .067 and .137 for option I and II, respectively). Participants consistently rated two arguments, TPL1M2 and TPL2P2, most important. Both arguments were related to the number of manual user entries. Consistent with interview data, these results indicate a strong disfavor toward manual interaction with the system.

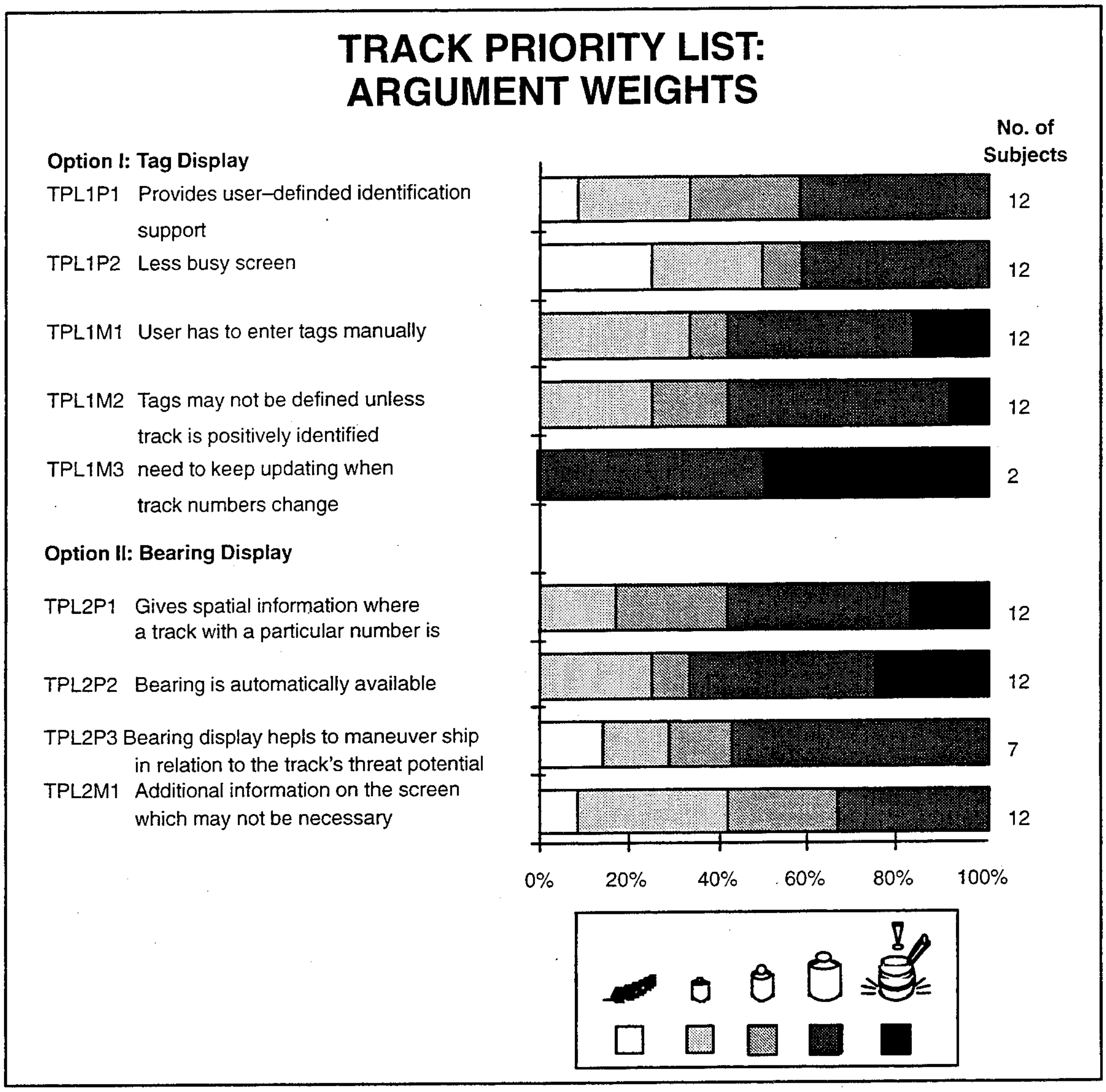

Figure 29. Proportions of participants' importance ratings for Track Priority List questionnaire arguments.

Researchers did not include TPL1M3 and TPL2P3 in the statistical analysis because of an insufficient number of observations. Note that TPL1M3, "need to keep updating when track numbers change," also relates to manual interaction with the system, which the two participants who responded to the item consistently rated important.

3.3 RESULTS FOR THE COMPLETE DSS

Researchers asked eleven participants to comment on the complete DSS screen. Dependent on the duration of the previous parts of the session (sometimes greater than 2.5 hours) and number of comments previously made, some participants had to end the session before researchers could gather this information.

Although all participants described highly individual styles regarding their interaction with the DSS, three groups of styles became apparent. There was a rather analytic approach with users looking first at analysis windows. The second, a "validating" approach consisted of picking up a track from the Track Priority List and verifying the system's priorization and response suggestions. The third group of users was reluctant to use the system at all or at least for the intended purpose.

Since all interaction patterns were different from one individual to another, results were impossible to summarize. Therefore, this report presents results for individual participants. It gives the individual selection of preferred window options and a schematic representation of the hypothetical interaction pattern with the system for every participant. The abbreviations used are AL for the Alerts Window, CN for the Comparison to Norms Window, SA for the SABER Window, RM for the Response Manager Window, and TPL for the Track Priority List. Arabic numbers designate options. AL2, for example, means Alerts Window option II. Figure 30 through 40 display schematic interaction graphs of participant's interaction patterns. The arrows in the graphs indicate the transition from one window (treats the ADS geoplot as a window) to the next.

3.3.1 Analysts Cluster

The analysts cluster consisted of five participants (5, 6, 8, 15, 16), thus forming the largest group. The interaction patterns described by these participants was the closest to the expected pattern (cf. overall DSS description in section I), using analysis windows first and then looking at priority and response issues. The group consisted of two TAOs and three COs. Note that four of the five members preferred the Comparison to Norms Window option I.

1. Participant 5 Preference: AL2, CN1, SA1 RM2a, TPL1 (figure 30). This participant said he would mostly look back and forth between the geoplot and the Track Priority List to see "what's important." If an alert occurred, he would look first at the Track Profile and the Comparison to Norms Window, selecting the track from the Track Priority List. If the track was a threat, he would consider the Template and SABER Windows. He felt the user had to become used to the Response Manager Window, but he found the time line useful. He would also use the SABER Window to explain to the CO why he had taken certain actions. He felt the SABER and Template windows were closely related because both deal with intent. He felt the busy screen was not a problem after becoming used to it.

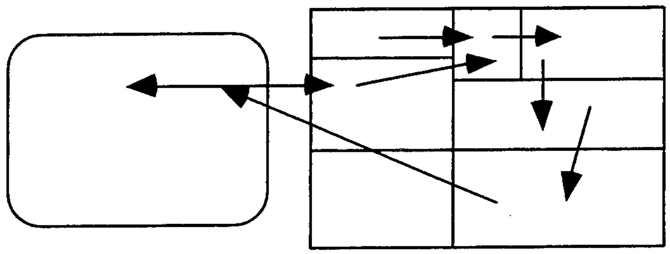

Figure 30. Participant 5 Preference: AL2, CN1, SA1 RM2a, TPL1 schematic interaction graph.

2. Participant 6 Preference: AL2, CN2a, SA1, RM2a, TPL1 (figure 31). This participant indicated he would mostly use the Alerts Window, Track Priority List, Track Profile and Comparison to Norms Window and keep responses and intent reasoning in mind. For a slow track like a helicopter, he would consider also using SABER. Once ID was clear he would ignore the Comparison to Norms information. The Alerts Window and Track Priority List would be used as a backup for memory and situational awareness. He pointed out that existing systems currently provide much of the Track Priority List's functionality. He considered the overall usefulness of the DSS "excellent."

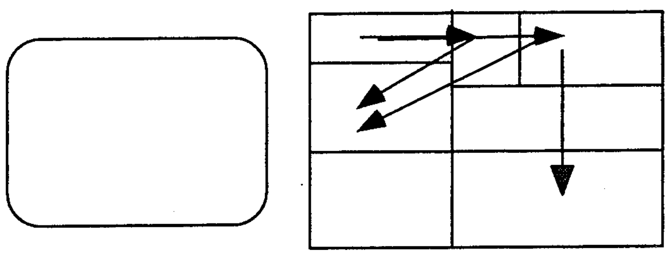

Figure 31. Participant 6 Preference: AL2, CN2a, SA1, RM2a, TPL1 schematic interaction graph.

3. Participant 8 Preference: AL2, CN1, SABER1, RM2a, TPL2 (figure 32). This participant indicated he would use the Alerts Window, the Comparison to Norms Window and the Response Manager Window "most of the time." In the case of a high speed inbound track, he would focus on Alerts and Response Manager. He would also use the SABER Window and "maybe" the Comparison to Norms Window for a slow helicopter because no quick action would be required. Helicopters were the most problematic tracks in the Persian Gulf. He found the Alerts, Comparison to Norms and Response Manager Windows "very good," and recommended making them available to other CIC workstations, possibly as an additional watch station. He considered the Track Priority List, SABER Window, and Track Profile also "useful." He felt the screen's complexity would be no problem because of future users' 6 months of training and hours on watch.

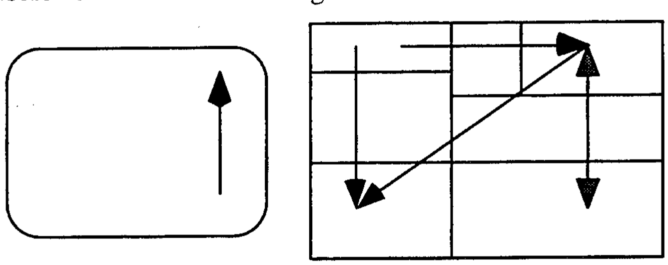

Figure 32. Participant 8 Preference: AL2, CN1, SABER1, RM2a, TPL2 schematic interaction graph.

4. Participant 15 Preference: AL2, CN1, SA1, RM2a, TPL1 (figure 33). This participant recommended that alerts have the track symbol flicker on the geoplot because the geoplot was his standard monitoring screen. If a track triggered an alert, which attracted his attention, he would look first at the Comparison to Norms Window, since his first question was about the ID of the track. For a track with unknown ID, the platform type was essential. The next questions were whether the track could shoot and with which weapons. Then he would look at its history on the Track Profile. Eventually, he would assign it a priority level and put it on the Track Priority List (if it was self-defined). After that, he would look at the Response Manager for responses under worst-case assumptions. If there was a problem, he would look at the rest of the screen. He felt that users would ignore the SABER and Template Windows most of the time. However, he felt a Template Window with an altitude over range picture would be very valuable.

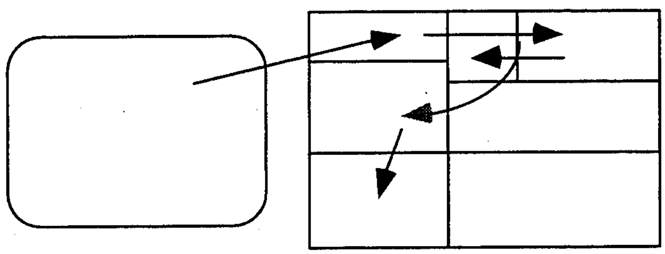

Figure 33. Participant 15 Preference: AL2, CN1, SA1, RM2a, TPL1 schematic interaction graph.

He suggested a different display principle. For three tracks (more could not be handled), he suggested displaying an integrated Template and Response Manager Window, based on an altitude over range display, for every track. The Track Priority List would update the selection of these three tracks. He suggested using the Comparison to Norms and SABER Windows as pop-up windows. He also suggested tuning the Comparison to Norms Window to the ID problem (variables: profile, country of origin, EW, IFF, ID maneuver, RTF profile, tactical air corridor). The advantage of three parallel displays would be that the user would not focus on one track only.

5. Participant 16 Preference: AL2, CN2, SA2, RM2a, TPL2 (figure 34). This participant indicated he would mostly use the Track Profile and the Comparison to Norms Window. He said he would also watch the Alerts Window. He considered the Track Priority List potentially "significant" and said he would check it periodically to make sure he was not focusing on the wrong track. Regarding this pattern, there would, at least initially, be no difference between high-speed and low-slow tracks.

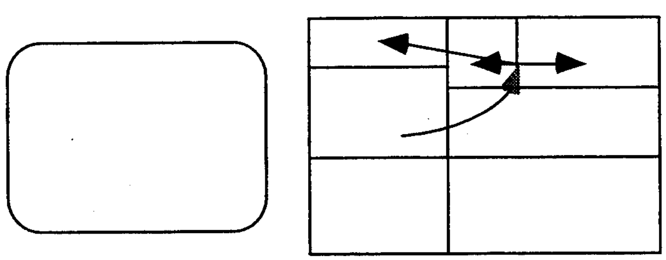

Figure 34. Participant 16 Preference: AL2, CN2, SA2, RM2a, Tpl2 schematic interaction graph.

He expected the Response Manager Window not to come up most of the time (although he found it "handy"). He would not use the Template and SABER Windows because he would not have enough

confidence in their intent reasoning. The Template display assumed that there were unique and identifiable contents for different hypotheses that might not be true. An altitude/range graphical template would be more meaningful than the version used here, but he still felt the Track Profile and Comparison to Norms Windows were more useful.

He also said that a Track Priority List is already available, and the Alerts Window and the Track Profile were mostly display technology.

3.3.2 Validaters Cluster

The validaters cluster consisted of three participants (9, 10, 12). Their approach was to look at the system's priority and response suggestions first, and then to check supporting and verifying information in the analysis windows. The group consisted of two TAOs and one CO.

1. Participant 9 Preference: AL1, CN1, SABER1, RM1a, TPL2 (figure 35). This participant expected color-coded, high-level alerts to draw his attention to the Alerts window. Next, he would "drop down" to the Track Priority List to look up the recommended next action. After that, he would look at the right side of the screen to see what supported this recommendation. The question was: "Does the system support my own decisions? Do we reach agreement?" In the high-speed track case, this participant indicated he might not look at the DSS at all. He would look at range, of course, IFF and immediately automate the close-in weapons system (CIWS). "The DSS is nice to have, but I would not rely on it. It won't take the place of training and experience." In the Iranian P-3 case in the TADMUS scenario B, he would first ask whether it was identified as Iranian. If so, he would ignore the Track Profile and the Comparison to Norms Window and immediately consider responses, using the Response Manager Window "as a cue for actions." He would use the SABER Window to look up evidence supporting his intent assumption. His experience said P–3s usually did not harass, but sometimes attack. If SABER Window displayed an unexpected hypothesis, he would be surprised and look up the listed evidence and the other analysis windows. All in all, he considered the Template Window "probably the least useful," and expected that he would usually go back and forth between Alerts Window and Track Priority List.

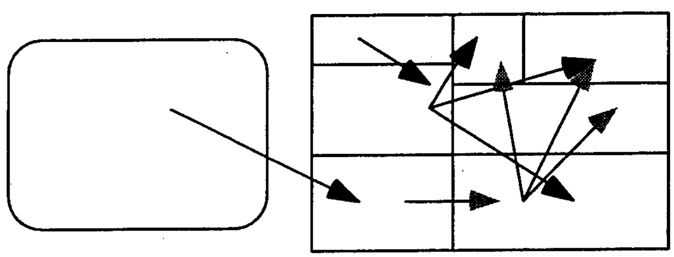

Figure 35. Participant 9 Preference: AL1, CN1, SABER1, RM1a, TPL2 schematic interaction graph.

2. Participant 9 Preference: AL1, CN1, SABER1, RM1a, TPL2 (figure 36). This participant indicated that he would not use the DSS if close to an engagement. In other cases, he would scan the Track Priority List (besides watching for alerts) to see "who to worry about next." Consequently, he would not look at the top line (if he had already worked on the top priority track) but below. Then, he would use the Comparison to Norms Window to validate the TPL information. He would farther use the Track Profile, Comparison to Norms, and SABER Window to validate the track's ID. Then he would "march through" the Response Manager, checking with the CIC team for what has already been done. The user would not click the option II feedback buttons until the system reported an action accomplishment (e.g., a warning issued). If busy, he would not press the feedback buttons.

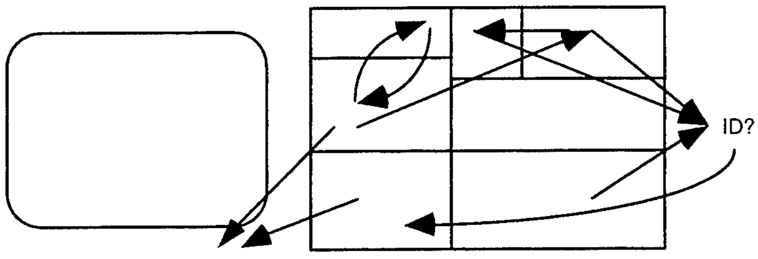

Figure 36. Participant 10 Preference: AL2, CN2a, SA1, RM2b, TPL2 schematic interaction graph.

This use pattern would be the same for slow helicopters and fast inbound tracks. Priority assessment would guide his use of the screen. He would farther use the Track Priority List and the Response Manager to brief the CO about the tactical situation and the state of the process. He recommended displaying intelligence information, indicating hostile intent more clearly. Currently, the system buries this kind of information in the SABER evidence lists. He also described himself as a "heavy EW user," trying to get as much information as possible from the EW operator, since he considered this information to be highly important.

3. Participant 12 Preference: AL2, TPF, CN2a, SA1, RM2a, TPL2 (figure 37). This participant said that the key window was the Track Priority List. After looking at the Track Priority List he would look down at the Response Manager Window to see what the response plan was for this track. Occasionally, he would quickly check the Alerts Window. The rest of the screen was "just available as a visual reference" (pointing at Track Profile). He considered the option II Comparison to Norms Window important because of its fusing of multiple parameters. Participants felt that the Template and SABER Windows provided useful background information explaining "why it is a priority track." In the case of a high-speed track, he would pick up the alert message and go right down to the Response Manager Window to look up responses. He would use the right half of the screen only as background information. In the case of a low, slow helicopter, he would try to determine the track's intent using the SABER Window.

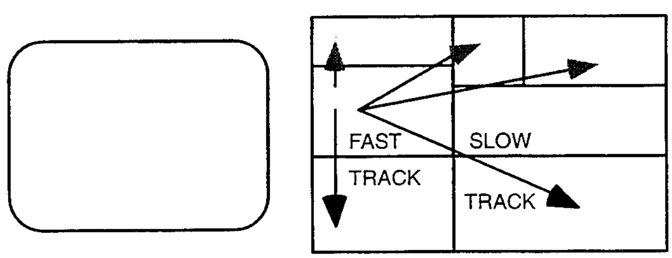

Figure 37. Participant 12 Preference: AL2, TPF, CN2a, SA1, RM2a, TPL2 schematic interaction graph.

3.3.3 Reluctant Users Cluster

The reluctant users cluster consisted of three participants (11, 13, 14). They thought they would hardly use the DSS at all, and only single windows. They reasoned that the DSS focused too much on one single track, distracting the user from the overall tactical picture. The group consisted of two COs and one TAO.

1. Participant 11 Preference: AL2, TPF, CN2c, TPL2, RM2b, no SABER and Template (figure 38). This participant felt that "verifying the machine's suggestions is crucial. People tend to rely too much on machines." Because he felt that users did not easily absorb the DSS' numerous alphanumeric displays, he suggested integrating the DSS logic (the RPD and SABER tools) in a three-dimensional display.

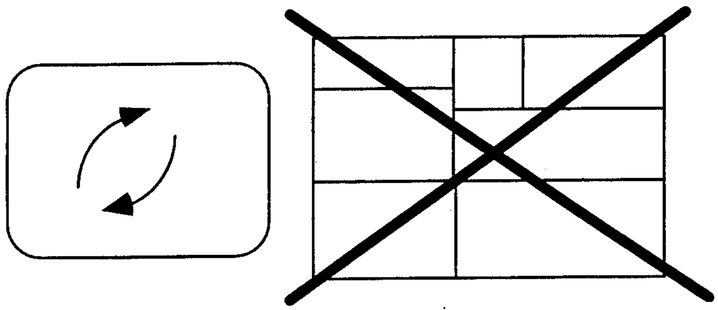

Figure 38. Participant 11 Preference: AL2, TPF, CN2c, TPL2, RM2b, no SABER and Template schematic interaction graph.

2. Participant 13 Preference: Alerts I, CN2c, SABERII, RMIIa, TPLII (figure 39). This participant indicated he would almost exclusively monitor the geoplot and use the DSS for amplifying information only. The geoplot showed the location of other friendly forces who could take care of a problem. Participants considered spatial relationships visible on the geoplot very important. If a track was threatening (inbound, unknown), he would use the DSS and/or information from a friendly craft closer to the track. His first look would be at the Track Priority List to verify if the system puts the threat on the first priority line. The CIC team should provide Track Profile and Comparison to Norms information. He would ask the team before consulting the DSS and balance and compare the information from the two sources. Users had to carefully evaluate the information's quality. Verbal reports might provide more recent information, which he considered to be crucial. The DSS would have the last priority except for the ROE support. He said that the accuracy of the DSS was not guaranteed, but that he was used to, and felt comfortable, with his CIC team's reports. Because he often had to deal with several tracks simultaneously, he was reluctant to use the DSS. The DSS would focus too much on one track. On the other hand, he would use the DSS more in the first 45 minutes of the watch to get more situational background. He felt the DSS would be more useful in a training context, and for less experienced users.

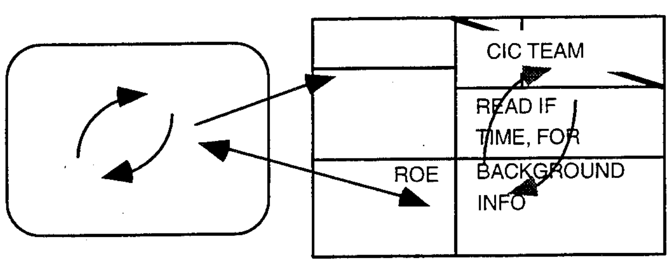

Figure 39. Participant 13 Preference: Alerts I, CN2c, SABERII, RMIIa, TPLII schematic interaction graph.

He suggested dedicating the whole screen to a few individually selected windows. He further recommended changing the arrangement of the windows according to their relative importance. He

would arrange the Track Priority List on the top left position, the Alerts Window right under that, and the Comparison to Norms Window right under the Alerts Window. On the right hand side he would arrange (from top to bottom) the Track Profile, the Response Manager, the Template and the SABER Windows. As a further reason for this arrangement, he explained that ROE were based on ID and relative threat, and that the Response Manager Window should be placed "after" the Comparison to Norms Window.

3. Participant 14 Preference: AL1, CN2a, SA1, RM2a, TPL2 (figure 40). This CO felt a CO should not manipulate keyboards and mice, as he was "the only one with the big picture." Consequently, he felt the CO did not need some of the windows, although other watch stations did, and these other watch stations should have access to integrated DSS functions. He would focus on the Track Priority List and talk to the TAO based on this to see whether there is an agreement. He further suggested switching the positions of the Alerts and the Track Priority List windows, because he considered the latter to be more important for the CO.

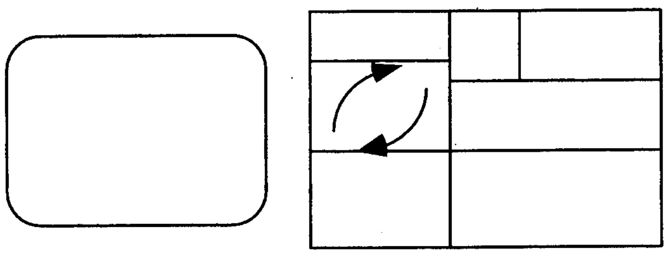

Figure 40. Participant 14 Preference: AL1, CN2a, SA1, RM2a, TPL2 schematic interaction graph.

4. DISCUSSION

4.1 OPTION PREFERENCES AND PRO/CON ARGUMENT WEIGHTS

The questionnaire data showed a consistent pattern, but pose a problem in summarizing the results: differences in subjective importance between questionnaire items were usually significant, but agreement among participants was low. Arguments that did not receive at least one "crucial" rating were very rare, and so were arguments that were not rated "negligible" by someone else. In the sections regarding individual windows, this report outlines the arguments that participants most consistently rated important. They are usually consistent with the interview data. However, note that experienced Aegis COs and TAOs also rated almost all other arguments "crucial," so they are not at all negligible. Thus, the following section, which summarizes the "highlights" in the data, is necessarily incomplete.

1. Alerts Window: Information should not be lost. Participants considered superseding alerts a serious problem in the option I display. Participants considered this argument, related to the windows even more important than arguments regarding the window's organization. Window modifications should provide retrieval functions or otherwise guarantee access to a comprehensive data set.

2. Comparison to Norms Window: A quick overview over many data sources. "Quick overview" arguments, indicating a low time demand for using a window, were consistently rated important and account for the discarding of option III. The number of accessible data sources was decisive for the preference for option II over option I, although the option II display was considered less clear and straightforward to use than the option I display. Modifications of the window should include multiple evidence factors and a display format that makes the window self-explanatory.

3. SABER Window: The system's suggestions must be verifiable. Participants consistently rated access to unfiltered evidence lists important because they considered the inability to verify the system's operation dangerous. However, this does not mean that the screen must display this information permanently.

4. Response Manager Window: A readable list including a feedback function. Participants considered option II's one-dimensional response arrangement sufficient. They considered greater legibility more important than option I's additional visual features. They also considered option II's feedback function important.

5. ROE Support: ROE support must be visible at any time, but not as a checklist. However, there was some concern that more complex ROE would not fit into the present study's small display box.

6. Track Priority List: Least possible interaction. Arguments rated most consistently important concern the fact that bearing was available automatically, whereas the user had to manually enter, and possibly re-enter when track numbers change, a track's tag.

Note that these principles governing participants' preferences are surprisingly unrelated to the displays' format and organization. These principles merely reflect a preference for more information instead of less information (Alerts, Comparison to Norms Window). They also show that the user needs the ability to verify the system's operation by accessing raw data (SABER Window). They show a disfavor of manual system interaction (Track Priority List). Issues related to display organization were only considered relevant in the Comparison to Norms and Response Manager Windows.

4.2 USEFULNESS RATINGS

Subjective usefulness ratings are not easy to interpret. The sign test is rather conservative in the current case because it does not differentiate between "useful" and "very useful" rating. This test clearly indicated all options' usefulness, except for the option III Comparison to Norms Window. Thus, except this window option, there was no justification for discarding any window from further experimental investigation.

However, an examination of average ratings shows that most single options received average ratings around 1 ("useful"). Successful options (with preference) like the Comparison to Norms Window sub-options 2a, b, and c received surprisingly low individual usefulness ratings.

Participants who considered one option for a window to be very useful, often rejected the other options completely. Researchers even observed this in the Track Priority List, where the options were only marginally different regarding their objective usefulness. Whether the screen displays the tag or the track's bearing should be unimportant to the priority list's principal usefulness. Apparently participants down-rated the non-preferred windows to underscore their preference decision. The theory of cognitive dissonance easily explains this phenomenon (Festinger, 1957, 1964).

To assess the usefulness of the underlying *concept* in each window, researchers considered the maximum rating given to the individual options more revealing than the average of all individual option ratings. The maximum rating probably over-estimated the concept's usefulness, but only slightly, while the average will underestimate it substantially because of the contrast effect discussed above. Figure 41 shows the maximum usefulness ratings for all seven windows.

Participants rated all windows, except for the Template and SABER Windows, equally favorable between "useful and "very useful" (the value of 1.6 is closer to the latter). They still considered the Template and SABER Windows useful, but less useful than the other windows. This probably reflects the opinion expressed by several participants that intent was less important than the researchers expected. Paragraph 5.3.1 discusses this issue and its impact on subjective usefulness in more detail. Another plausible drawback regarding the subjective usefulness of these windows is their alphanumeric format, which, on the other hand, was well accepted for the Response Manager Window.

4.3 INTERVIEW DATA

4.3.1 Participants' Comments

This report does not discuss all individual comments and remarks in detail. Refer to Appendix B for a complete list of comments. The comments' most prominent feature is their individuality and diversity. Only one or two participants brought up most issues, although other interview sessions discussed some of these issues implicitly. It is not appropriate to assign a relative importance to these comments simply by the number of participants who raised the issue. The more important question is whether researchers can learn from them, which obviously requires a subjective answer.

The majority of participants disliked alphanumeric displays and favored colored graphics and pictures, even if there was a nominal loss of information content and flexibility. As SABER Window comments underline, many participants felt they had to read through all alphanumeric information presented before they could decide what was important. Also, it is likely they would not accept higher level, integrated information (assumptions, intent hypothesis suggestions) without knowing the information background (i.e., data sources and integration algorithm). So participants generally

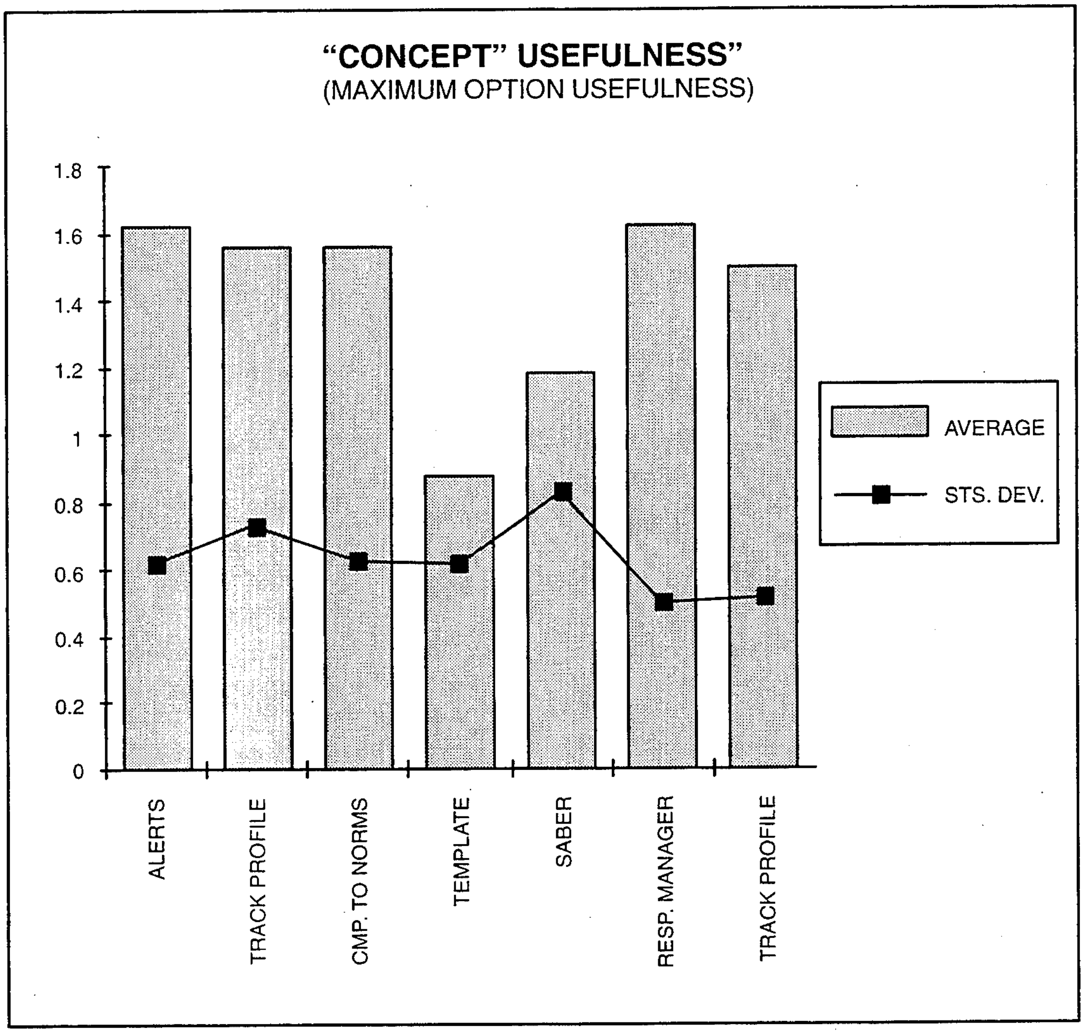

Figure 41. Maximum usefulness ratings for the respective options of the seven DSS windows. Averages and standard deviations were calculated over participants.

preferred graphical displays because they perceived that the filtering process to extract relevant information was more time-consuming in the alphanumeric displays than in the graphs.

Most participants also disliked clicking buttons. Some participants even felt that current systems already made too high a demand regarding console functions. Therefore, in developing the DSS, researchers should keep interactivity to a minimum. On the other hand, several participants requested personal editing and set-up functions for use during the mission preparation. This means developing generic editing and set-up tools that users can easily and flexibly use, without susceptibility to contradictory or inappropriate settings.

A counter-intuitive finding expressed by several participants was that range is more appropriate than time as a base for event sequence displays (Track Profile, Template, and Response Manager Windows). From a task-analytical viewpoint, researchers expect that time is a more common base than range. Users have to make mental calculations to determine the exact timing for responses or track behavior expectations if the display provides only the track's speed and range. However, they are trained and used to making these calculations, so the costs of this particular processing step are

probably negligible. There may be two additional explanations regarding the preference for range-based displays. First, tactical manuals provide attack profile graphs based on range scales, so users are familiar with this display format. Second, as participants pointed out, users base tactical reasoning on range data like weapons release ranges, battle order thresholds, etc., while time information may vary with every case. Basing templates and response plans on a range scale rather than a time scale makes a more general and recognizable display. Displaying 1-minute speed leaders still supports integrating the range and speed information into time interval estimates. The speed leaders indicate the range interval the track will travel within the next minute if its speed remains constant.

Like most complex systems, users can use the DSS in several unexpected ways. Users can use the SABER and Response Manager Windows, as well as the Track Priority List, as briefing tools. These tools inform the CO about the tactical situation when the CO arrives in the CIC. This may indicate that they help understanding the tactical situation. Other unexpected, but possible uses indicate that it will be hard to keep some windows functionally separate. The option I Comparison to Norms Window, e.g., displays information about a track's kinematic history, which is the principal function of the Track Profile. Users can use the Alerts Window, especially option II, to look up tactically significant events in a time-stamped window. This would give the Alerts Window part of the Template and SABER Windows' functionality. Researchers only partially intended this functional overlap between windows. Since there is a hierarchy of integration levels between the windows, a certain overlap in the information content of windows is unavoidable. However, researchers can only assess an individual window's additional impact on the decision making process when compared to another window, if the screen displays the information in both windows in a comparable format. Otherwise, the format and content factors would be confounded. In the ongoing HCI development process, researchers have to identify such functional and informational overlaps to provide comparable display formats. This discussion also showed that an integration of DSS and Aegis display system (ADS) functions would be useful and necessary in an applied system, but may decrease the separation of experimental factors. Researchers should not attempt this integration in the experimental system. However, the interview data show a strong need for integration in any future DSS applied version.

Contrary to previous research (e.g., Zachary et al., 1992; Miller et al., 1992), several participants indicated that in ambiguous littoral scenarios they would spend more effort determining a track's ID than its intent. Furthermore, they felt intent was so difficult to determine that they would not base tactical decisions on intent assumptions (and even less on a computer's). Some participants agreed that they would think about intent in certain low and slow tracks like helicopters, but certainly not in every case. If researchers assume that intent is not the question a user has in mind, researchers could still use the Template and SABER Windows as event and evidence lists. Participants spoke almost exclusively of the SABER Window as an evidence list. Comments on the Template Window's tactical use show that participants barely spoke of intent, but rather worried whether a track *could* attack. Subjective usefulness data supports this hypothesis. If users merely use the Template and SABER Windows as events and evidence lists, their subjective usefulness should be less than the usefulness of the Alerts Window (option II). The Alerts history function also provides a comprehensive, time-stamped list of tactically significant events on a single track, and a warning function that provides additional functionality. Indeed, all participants but one (14) assigned a higher usefulness to the option II Alerts Window than to the Template Window. The disagreeing participant, a CO, disliked any Alerts Window because he felt it drew his attention away from the global picture (which the Template does not).

An unanticipated finding was that many participants indicated they would very much appreciate a smart, context-sensitive ROE support system. Such a system would display key ROE statements,

based on the tactical situation, the track's platform type, country of origin and behavior, in a small window close to the response plan. However, system development for complex ROE is beyond the scope of the TADMUS program.

4.3.2 Complete DSS Use Styles

The most prominent feature of the interview data regarding the envisioned use of the complete DSS is the heterogeneity of opinions. There was virtually no matching pair of interaction patterns, and still much heterogeneity was present within the three-user style groups. There are three possible explanations for this finding.

First, the decision making process itself can be substantially different between individual tacticians. Interview data on the Track Priority List show that participants suggested several different principles for the priorization of tracks. Participants also indicated different variable sets for track identification. Further inter-individual differences exist in the details an officer chooses to process: some participants emphasized the "big picture" while others expressed interest in a single track's detailed information.

Second, the DSS structure may not match the mental representation of the task domain. The human-computer interface investigated here displayed seven independent windows, whereas the user has to integrate all this information into one tactical picture. The window's functional separation may seem artificial to a user. Participants recommended window integration several times. If this separation is artificial, researchers cannot expect users to re-integrate the tactical picture universally.

Third, participants lacked experience using the system. Researchers based this report's results on subjective data collected on static pictures. Patterns of user-system interaction can become more consistent when users have more opportunities to gather experience with an operational DSS.

The three-user style categories set is logically exhaustive. If the user uses the system at all (i.e., a user is not reluctant), information processing can be bottom-up (analytic) or top-down (validating). Both are reasonable approaches to using the system. The analytic style is closest to the general approach taken in the TADMUS program, i.e., to build a system that supports the user in making decisions by providing appropriate information elements. The validating style, on the other hand, makes full use of the system's decision making capabilities, and processes detailed information only if there is time or an obvious need to do so. Whether a user chooses a validating or analytic approach may depend on various factors like trust in the system, time available, ambiguity of the situation, etc. Users have to determine these factors empirically. It is conceivable that participants in the present study envisioned different tactical situations and came to different conclusions on the approach they would take.

A closer examination of the reluctant users' interview data shows that they are far less negative than it appears. One participant (11) explicitly stated that "verifying the machine's suggestions is crucial" (a "validater's" view). He also disliked the alphanumeric format of most windows, favoring a three-dimensional, graphical display of the features used in the investigated DSS. Another participant (13) demonstrated the attitude of an experienced TAO accustomed to a well-managed CIC team. There is no doubt that a good CIC team does much of the DSS's job. However, this participant would use the DSS to *understand* a tactical situation. This is not far from the intended use of the DSS as a support, not a substitute for human decision making. The third "reluctant user" (14) finally liked most of the windows and suggested making them available to other CIC watch stations.

The kernel of the "reluctant users" criticism of the DSS was that the DSS focused too much on one track, which might distract a user from the global tactical picture. Some non-reluctant participants also considered this a problem. Furthermore, the DSS's logic is necessarily egocentric. The DSS ignored the presence of and possible support by friendly forces. It also ignored the impact of other objects close to ownship (e.g., being in the way of an attacker, or being themselves the targets of an attack). While researchers can attack and empirically investigate the first problem by displaying selected analysis windows for multiple tracks, the latter problem is partially an artificiality of the experimental scenarios. However, researchers must consider this in future developments.

4.3.3 Participants' Recommendations

Participants have made several recommendations. Appendix C lists them with some comments. Participants' individual recommendations reflect abundant experience in computerized tactical console use. However, since the design goal for an experimental DSS (maximum diagnosticity regarding display elements) is different from the design goal for an applied system (maximum usability), researchers cannot immediately follow most of them. They will postpone them to a later development stage.

The following chapter gives an annotated overview on the most important recommendations.

Several recommendations regarded window integration. For example, researchers could indicate alerts by flickering symbols on the ADS geoplot. The user could click on the symbol to get the DSS information. This would spare a separate DSS trackball, but it is technically impossible in the current DEFTT system configuration.

For the Alerts Window, participants made several recommendations to increase the window's diagnostic value, e.g., more specific messages, the history function as a default, better highlighting of selected tracks, etc. These are useful modifications, but they add functions to the Alerts Window beyond the user's attention towards a track. Participants recommended allowing the user to edit the tripwires' triggering alerts. This recommendation would be useful in applied systems, but would spoil experimental standardization in the empirical validation process.

Participants recommended altitude over range and polar range/bearing history displays for the Track Profile. These recommendations do not interfere with experimental goals and researchers can follow them. Further recommendations to add weapons release ranges and similar information would confound the window's content with the Template and the Response Manager Windows.

Several participants recommended adjusting scale ranges to accommodate for the full range of possible values, e.g., of altitude. Researchers have to evaluate these recommendations considering a fairly constant mapping of the track profile to support recognitional processing.

For the Comparison to Norms Window, several participants made recommendations regarding the platform type and variable sets. They favored an even more multi-dimensional display. For the color set, they recommended more saturated colors several times. On the other hand, researchers must define the color set considering perceptual issues to maximize the platform type's detectability without misfitting data. Saturated colors can interfere with the suggestion that green and yellow matrix cells both indicate compatibility with a hypothesis. In addition, saturated colors would suggest a warning status.

Several participants recommended inclining the vertical text lines. Researchers tried this measure before the evaluation study. They found that it does not increase legibility on the monitor's rectangular pixel array.

For the Template Window, participants recommended basing the window on a range scale instead of a time scale. Researchers accepted this recommendation because of the reasons indicated in paragraph 4.2.4.2.

For the SABER Window, participants recommended both select and differentiated filtering to reduce the text items displayed. However, participants emphasized (correctly) that the user must know what is filtered out of the list. The preference data show that participants considered no filtering at all the best default.

For the Response Manager Window, automated feedback was recommended several times, which is consistent with all other data sources. Consistent with the Template Window and the Track Profile Window, participants recommended a range scale-based display. This recommendation can be accepted, although there is a conflict between the reading direction (left to right) and the items' arrangement on a range scale (range increasing to the right, arranging items in sequence from right to left). There were some controversial recommendations regarding the window's content, which must be user-defined in an applied system, and indicate important items (supposing the existence of unimportant ones).

Participants appreciated the ROE Support box format. They recommended making it context-sensitive. They also wanted to cancel the checkmarks indicating accomplished ROE criteria. Researchers can follow both recommendations without interfering with experimental goals.

For the Track Priority List, several recommendations were made regarding the priorization logic. Since these differed from each other, here again a user-specific approach might be beneficial in applied systems. One participant even suggested defining track priority by manual selection of priority tracks. The functional similarity of the Track Priority List and the Alerts Window (both directing attention) was reflected in recommendations to integrate both, displaying alerts messages within the Priority List. This would be possible even within a controlled experimental plan using an alerts history function like in the investigated option II display, but it would make the Track Priority List even more complex and "busy."

Recommendations for the Complete DSS Screen addressed two important issues: making DSS functions available for multiple tracks at a time, and transferring DSS functions to other CIC operators. The usefulness of both recommendations strongly depends on personal working styles and the CIC organization a CO/TAO has in mind. A user might prefer access to detailed information at any time, or avoid being distracted from the "big picture" and delegate analysis tasks. Both positions have merit depending on the management philosophy one follows. It is important to point out that individual differences exist with major implications for the layout of tactical DSSs.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 DISPLAY PRINCIPLES FOUND

While researchers based the present investigation solely on subjective data, the results outline principles for future applications that use the proposed TADMUS DSS concepts. Participants accepted these concepts very well and considered them useful. However, due to the heterogeneity of opinions, researchers expect that individual users will remain skeptical about single display concepts, dependent on their personal style. Researchers will not find a single, representative opinion.

A CO (14) outlined a possible solution to this problem. He indicated that he would not use the presented DSS himself, but strongly recommended making its functions available to certain CIC operators. As COs organize their CIC teams in different ways, they might also individually prefer to rely more on human operators who could use parts of the described DSS.

Participants considered the SABER and Template Windows more controversial than other windows, even though they considered them useful. The reason for this is not the display principle or the underlying logic, but several officers' opinion that a track's intent is less important than its ID. Note that all participants did not share this opinion, either. However, this is not a problem of display principles, but of understanding the task domain.

This section discusses principles for the seven individual windows, representing the display concepts investigated.

The Alerts Window confirmed some previously known principles. One confirmed principle was that researchers should avoid inflation of alerts (first described by Veitengruber, 1978; see Wickens, 1984, for instructive examples). Another was that alerts should not interfere with the use of other console functions if they do not require an immediate response (*caution* or *advisory* signals, according to Boff & Lincoln, 1988). The present investigation's results show that previous alert messages must be easily accessible, since they may help analyze the situation by providing a list of significant events. Whether the screen flags a single track's events or displays them in a coherent list are less important than the previous alerts' actual availability. The number of alerts that should be visible in an Alerts Window may depend on the situation, but participants accepted the proposed window size (six alerts). Dependent on the cognitive structure of the task domain, the "magical number" 7 ± 2 (Miller, 1956) seems to be a good guess. Participants also recommended that users control which events trigger an alert. They suggested that users personally edit hierarchical tripwires.

Researchers intended the Track Profile Window as a time history display of important system parameters. However, participants viewed it in a slightly different way. Participants indicated they preferred a display based on a range scale instead of a time scale, because this is a more generic format and a more appropriate representation of tactical reasoning. The key variable in the present task domain is altitude, whereas specific range and speed history are less important than trends. So one participant suggested a polar display because range history is important in the context of spatial relationships.

An altitude over range display provides the opportunity to add overlays (engagement ranges, ROE thresholds, tactical tripwires), which would add much of the functionality of a Response Manager or Template window.

The usefulness of the Comparison to Norms Window depends on the proper definition of norms. If norm boundaries are fuzzy or non-diagnostic, it loses value. Participants were aware of this fact

and preferred multi-variable displays over less-dimensional displays that might be easier to use. However, there was the concern that such a display might lead the decision maker to an incorrect assessment, e.g., by discarding a possibility because of unreliable data, especially in fuzzy norm boundaries.

A multi-variable display requires a coding format to indicate an object's features fit or misfit in the variables' specific norm boundaries. For this type of coding, color was subjectively preferred over fill patterns and matrix cell shape variations.

In the Template and SABER Windows, many participants disliked the alphanumeric format, even though it might be more flexible in presenting information on different hypotheses. In the present context of tactical reasoning, participants preferred arranging Template display events on a range scale instead of a time scale. They expected this arrangement to make the templates more consistent and recognizable. Several participants considered it problematic in both displays to devote the window to the problem of a track's intent. They felt that it was very hard to determine whether a track intends to harass or attack. However, this is more a problem with the window's content than with its display principle. Researchers may use the SABER Window as a briefing tool as well as a tool to verify the system's suggestions and proper operation. This is because it holds (and, according to participants' preference, is supposed to hold) a comprehensive list about a track's current evidence.

For the Response Manager Window, one scale (range or time) suffices for action arrangement. Participants accepted a time scale, but a range scale displaying the track's actual range and a 1-minute speed leader will do as well and be more consistent with the Template Window. Researchers can use the same symbology to save training time. Participants considered feedback on actions already begun and/or accomplished important. Researchers base feedback on automated messages from the effector side due to the possibility of a high rate of actions to be performed. Given an appropriate feedback function, participants considered the Response Manager Window a very useful briefing tool that may improve communication speed and accuracy. Users could effectively inform the CO, who will not always be present in the CIC in low-conflict scenarios, upon arrival in the CIC.

Participants considered a ROE Support tool helpful in the context of (but not integrated within) the Response Manager Window. It should not be in a checklist format to avoid premature engagement. Since actual ROE are much more complex and less concrete than the experimental ROE used in the TADMUS scenarios, researchers would have to distill and filter using context-sensitive and platform-sensitive display rules.

A Track Priority List already exists in the Aegis environment, but its prioritizing logic and display update delay limits its usefulness. In ambiguous scenarios, the currently used priorization rules (range-by-range rate, last opportunity to engage) are considered sub-optimal. Participants ordering the list by a combination of ID (hierarchy: hostile > UAE > UUE > UAF), time until the track reaches its weapons release range, last opportunity to engage, and possibly other factors like profile, region, and political situation. Researchers should adjust the tradeoff between the number of factors considered and the processing time so that top priority tracks are ordered within an accuracy of one to two ranks. Participants referred to information in the investigated Track Priority List displays other than the priority ranking, as possible briefing data that might helpfully explain the tactical situation. However, participants suggested an integration with the Alerts Window (alert messages displayed *in* the priority list), and appreciated displaying each track's bearing and range as an orientation support.

5.2 EXPERIMENTAL DSS

This investigation indicates that researchers can use apparently simple displays in several different ways, and that secondary emergent features may be unintentionally present. For example, the Comparison to Norms Window (option I) showed speed and altitude of a track, speed and altitude boundaries for various platform types, and the track's speed/altitude history, in a two-dimensional graph. The track's history "tail" turned out to provide information on the track's kinematic profile, i.e. speed and altitude history. Several participants recognized this information spontaneously.

For the sake of factorial separation in the TADMUS experimental plan, which will be necessary to determine every window's impact on user performance, researchers must carefully control such side effects. Since they are, on the other hand, beneficial to the user, they are also welcome to a certain extent. As long as the amount of informational overlap between windows is small and well-defined, statistical control of the resultant effects will be sufficient. If two windows share a display element, this would result in a two-way interaction effect. However, to avoid misinterpretation, researchers must know exactly which display element is concerned.

The seven windows investigated here present features of the tactical situation at different levels of integration and in varying contexts. Thus, some overlap necessarily occurs in the window's contents. Content overlap exists for all seven windows. Controlling the display format of these overlapping contents over the respective windows is crucial to ensure a clear separation of display factors that influence performance. Otherwise, the impact of a window's information content would be confounded with the impact of its display format. Alternatively, researchers could display common in two windows in a constant format, or in such different formats and contexts, that a user would very unlikely extract information spontaneously from the "wrong" display element.

For example, in the Alerts, Template and SABER Windows, present lists of tactically significant events. All three windows present this information in the same format (i.e., alphanumeric), so researchers can successfully avoid format/content confoundation. The Track Profile and the Comparison to Norms Window both display kinematic data, and the option I Comparison to Norms Window risks confounding information content and display format. However, the option II Comparison to Norms Window is so different from the Track Profile that the risk of confoundation is low. It would be implausible to expect a user to extract profile information from the option II Comparison to Norms Window or classify information (at high resolution) from the Track Profile.

The high variability in user strategies and styles requires attention to experimental control. It will be necessary to either control, or monitor the user's access to information. Researchers can accomplish this by exerting tight control on the information displayed, or by monitoring the user's information gathering behavior. Considering the rather moderate resolution of current eyeball tracking systems (1 to 2 inches maximally), using seven separate windows is a viable option. Other approaches, such as having users actively ask for information (e.g., by clicking buttons to have it displayed), are possible, but they also interfere with participants' natural behavior and preferences.

5.3 FUTURE DSS VERSIONS

The results reported here yield questions and hypotheses for future research. Since several participants indicated that the DSS focuses too much on a single track, researchers could investigate alternative DSS versions based on selected windows. Researchers could base selections on objective evaluation experiments, but display information on multiple tracks. Researchers could use personally configurable displays to further investigate the interaction of user strategies and preferences with DSS configurations.

Participants strongly recommended personally editable tripwires, norms, templates and action plans. Researchers will have to build the respective editing and configuring functions. Researchers will have to use advanced assisting tools to avoid contradictions and inappropriate settings.

The transition into an applied system will require several changes to the DSS design. Researchers will have to adapt the color set to the current fleet lighting and operating conditions. Since there is no longer a need for factorial separation, researchers can integrate useful windows, and graphical displays can use more extensively. Furthermore, the system will have to deal with variable data reliability. The future DSS should display the respective information's credibility and the evidence display. Finally, researchers will need to address the problem of changing track numbers. If the system loses a track is momentarily and it appears again under a different track number, the system must easily transfer DSS information to the new track number.

Obviously these problems go beyond the scope of this research document. However, researchers have established a foundation for display principles.

6. REFERENCES

Bennett, K. B., M. L. Toms, and D. D. Woods. 1993. "Emergent Features and Graphical Elements: Designing More Effective Configurational Displays," *Human Factors, vol. 35 (1),* pp. 71–97.

Boff, K. R., and J. E. Lincoln. 1988. *Engineering Data Compendium.* Harry G. Armstrong Aerospace Medical Research Laboratory, Wright-Patterson AFB, Ohio.

Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Harper and Row, New York.

Festinger, L. 1964. *Conflict, Decision, and Dissonance*. Stanford University Press, Stanford, CA.

Hair, D. C., and K. Pickslay. 1992. "Explanation-Based Reasoning in Decision Support Systems." Proceedings of the 9th Annual Conference on Command and Control Decision Aids. June, Monterey, CA.

Hockey, G. R. 1986. "Changes in Operator efficiency as a Function of Environmental Stress, Fatigue, and Circadian Rhythms." In: *Handbook of Perception and Human Performance.* K. R. Boff, L. Kaufman, J. P. Thomas, Eds. Wiley, New York.

Hutchins, S. G. 1995. "Decision Making Evaluation Facility for Tactical Teams," Naval Postgraduate School Technical Report. In Preparation.

Hutchins, S. G., and J. T. Kowalski. 1993. "Tactical Decision Making Under Stress: Preliminary Results and Lessons Learned." Proceedings of the 10th annual Conference on Command and Control Decision Aids. June, Washington, D.C.

Kaempf, G. L., S. P. Wolf, M. L. Thordsen, and G. A. Klein. 1992. "Decisionmaking in the AEGIS Combat Information Center," NRaD Technical Report,Task 1, San Diego, CA.

Klein, G. A. 1991. "Recognition-Primed Decisions." In *Advances in Man-Machine Systems Research, vol. 5*, pp. 47–92. W. Rouse, Ed. JAI Press, Inc., Greenwich, CT.

Klein, G. A. 1992a. "Decisionmaking in Complex Military Environments," NRaD Technical Report, Task 4, San Diego, CA.

Klein, G. A. 1992b. "Summary of Data Collection and Results for TADMUS," NRaD Technical Report, San Diego, CA.

Miller, G. A. 1956. "The Magical Number Seven, Plus or Minus Two," *Psychological Review, vol. 63*, pp. 87–97.

Miller, T. E., M. L. Thordsen, S. P. Wolf, and G. A. Klein. 1992. "A Decision-Centered Approach to Storyboarding Anti-Air Warfare Interfaces," NRaD Technical Report, San Diego, CA.

OSF/Motif. 1990. *Style Guide*. Prentice Hall, Englewood Cliffs, N.J.

Pomerantz, J. R. 1986. "Visual Form Perception: An Overview." In *Pattern Recognition by Humans and Machines, vol. 2,* pp. 31–61. H. C. Nusbaum and E. C. Schwab, Eds., Academic Press, Orlando, FL.

Sanderson, P. M., I. Haskell, and J. M. Flach. 1992. "The Complex Role of Perceptual Organization in Visual Display Design Theory," *Ergonomics, vol. 35 (10),* pp. 1199–1219.

Smith, D. E., and J. Grossman. In preparation. "Understanding and Aiding Decision Making in Time-Constrained and Ambiguous Situations," Submitted for publication at NRaD, San Diego, CA.

Veitengruber, J. E. 1978. "Design Criteria for Aircraft Warning, Caution, and Advisory Alerting Systems," *Journal of Aircraft, vol. 15*, pp. 574–581.

Wickens, C. D. 1984. *Engineering Psychology and Human Performance*. Charles E. Merrill, Columbus, OH.

Woodson, W. E. 1981. *Human Factors Design Handbook*. New York: Mc Graw-Hill.

Zachary, W. W., A. L. Zaklad, J. H. Hicinbothom, J. M. Ryder, J. A. Purcell, and R. J. Wherry. 1992. "COGNET Representation of Tactical Decision-Making in Ship-Based Anti-Air Warfare," CHI Systems, Technical Report 911015.90109, Springhouse, PA.

7. GLOSSARY

ADS	Aegis Display System
AL	Alerts Window
CIC	Combat Information Center
CIWS	Close-In Weapons System
CN	Comparison to Norms Window
CO	Commanding Officers
CPA	Closest Point of Approach
DEFTT	Decision Making Evaluation Facility for Tactical Teams
DSS	Decision Support System
EW	Electronic Warfare
HCI	Human-computer Interface
ID	Identification
IFF	Identification Friend or Foe
NCTR	Non-Cooperating Target Recognition
NTU	New Threat Upgrade
RM	Response Manager
ROE	Rules Of Engagement
RTF	Return To Forces
SA	SABER Window
SABER	Situation Assessment by Explanation-Based Reasoning
TADMUS	Tactical Decision Making Under Stress
TAO	Tactical Action Officers
TPF	Track Profile
TPL	Track Priority List

APPENDIX A: PROCEDURE SCRIPT

INTRODUCTION TO EVALUATION EXPERIMENT

Motivation

"As the TADMUS Decision Support System (DSS) is being designed as an experimental system in an experimental environment, we are operating in a "standard-free space". Thus, we have some additional degrees of freedom in designing the human-computer interface. Furthermore, we have some design alternatives where our "real-task" knowledge is just not sufficient to decide upon. We would like to let you participate in the development process in order to take advantage of your operational knowledge.

Since this is an experimental environment, what you will see here is quite different from what we might eventually implement aboard. Anyway, we want to make sure at a very early point of the development process to eliminate the useless stuff and design as close to your needs as possible."

Schedule

"As a warm-up, we will show you a video sequence of your own performance in one TADMUS experimental scenario. We'd like to encourage you to comment your thoughts and actions in the scenario. The purpose of the video sequence is to bring you back into the thinking process both from the "inside" view and the view "from above" the tactical situation in the experiment.

After the video sequence, we will present the various tools or components that make up the Decision Support System. The purpose of the system, the function and the arrangement of the windows on the screen will be explained. In the meantime, the scenario shown in the video sequence will be started and frozen at an early point to give a visual cue about the operational situation.

Then, we will talk through all DSS windows and present alternative design options. Where alternative options exist, we will ask you to state your preferences and fill out a questionnaire about the relative weight of arguments that can be used for the decisions. You are welcome to add any number of additional arguments you consider to be important.

Next, we will ask you to tell us how useful you find the different windows (including alternatives) by sorting corresponding file cards onto three piles (very useful, useful, and useless).

While you are sorting cards, we will build a full DSS screen using your preferred window options. We then would like to discuss how you would use this version of the DSS. To support realism, we will run a TADMUS scenario at the same time."

ANY QUESTIONS?

"We hope you will enjoy the experiment."

[Demographics collection if not yet done]

VIDEO REVIEW

[Presentation of a scenario portion where a benefit from the DSS is most likely to be seen]

INTRODUCTION TO DSS

[DEFTT scenario is started and frozen after all tracks appeared]

[DSS slide show: title]

Purpose of DSS and its windows. "The DSS is designed to support a natural decision making process. We try to present raw and integrated data in a way our previous research has shown it is likely to be used by human decision makers. Thus, the DSS does not make decisions of its own, but it tries to assist you in certain stages of the decision making process."

"The DSS currently consists of the following seven windows: Alerts Window, Track Profile, Comparison to Norms Window, Template Window, Intent Hypothesis Window, Response Manager Window, and Track Priority List."

Purpose of current HCI. [Slide: Window Arrangement]

"The current design of the human-computer interface is tailored for experimental purposes. The seven windows are all tested as stand-alone applications as well as in a complete system configuration to determine the influence of each window on the decision making process. Therefore, some redundancies could not be avoided."

Arrangement of windows in accordance to their purpose. Alerts Window: provides comparison of evidence of events to thresholds.

Track Profile: presents time history of a variable (altitude, range, speed) as an explicit feature.

Comparison to Norms Window: provides a quick comparison of features for one contact to features for exemplars of other contacts.

Template Window: Assembles lower level features and compares them against reference values. Relates individual events, presents hypothesis for the situation based on integration of events.

SABER window: displays causal relationships between individual events, presents hypotheses for situation, based on causal model, presents evidence for hypothesis, presents assumptions required to accept hypothesis.

Response Manager Window: provides assistance in using preplanned responses.

Track Priority List: provides integrated picture: ID, intent, priority and why, next action to be taken (preplanned response), time to take next action.

"The windows are arranged in order to ensure a continuous working cycle. Shortcuts can be taken from whatever window to another one."

Some General Conventions

Track information is white, system-generated information is light blue.

All active screen elements, such as click-buttons and entry fields, are bluish gray rounded-corner rectangles.

Selected click-buttons are highlighted.

SELECTION PROCESS

Alerts Window

Introduction Alerts Window. [Slide: Alerts option I]

Purpose. "The Alerts Window provides a comparison of evidence of events to pre-defined thresholds."

Description. "There are two options for the Alerts Window.

Option I consists of a list of alerts, ordered by time. Since the RAINFORM GOLD messages give an update every 6 s there may be several alerts arriving at the same time. These alerts are ordered by increasing track number.

Each alert is displayed on a line with the corresponding track symbol, the track number button, the time of the alert, and a cancellation button. Clicking on the cancellation button will eliminate the respective alert.

Clicking on a track number button as well as selecting this track for analysis (whichever way) in another window causes highlighting of all buttons displaying the same track number (option I only), so that all alerts about this track can easily be discerned.

The alerts are colour-coded corresponding to the current air warning colour code.

New alerts flicker for 2s to gather attention."

[Slide: Alerts option II]

"Option II tries to avoid a possible confusion of track numbers. In this option, only the latest alert per track is displayed (the most recent alert still in the top line). By clicking on a "History" button, earlier alerts about this same track are displayed in a window popping up right under the line under consideration. Previously canceled alerts are designated in this window by checkmarks. You can return to the previous display by clicking a "Return to list" button or by specifying another track for analysis.

Note that in the option II Alerts Window the history of a previously canceled alert can not be displayed since the alert line with the "History" button is no longer present. Only when a new alert occurs regarding the same track the alert history is again available. Since there will be only one alert per track displayed and thus more room will be available, the cancellation button has a less important function in this display than in the option I display."

ANY QUESTIONS?

Presentation of the questionnaire. "This questionnaire is designed to give us an idea about why you prefer a certain option. The questionnaire consists of a list of arguments in favor of or against every option. We want to know how important you consider the different arguments to be.

Please check the appropriate symbols according to your opinion, but feel free to make your decision independently from that (actually, we conduct separate analyses on the questionnaire and your preference choices). Feel free to add any arguments we might have overlooked."

Selection option I vs. option II. "Do you prefer the "plain list" alternative or the "pop-up history" alternative?"

Further recommendations?

Track Profile Window

Introduction Track Profile Window (no selection). [Slides: Track Profile]

Purpose. "The Track Profile presents the time history of a variable (altitude, range, speed) as an explicit feature."

Description. "The Track Profile window displays time history graphs of selectable variables. Those variables include altitude, speed and range for air tracks, and altitude and speed for surface and sub tracks. The variable to be displayed is selected by clicking on the respective button. Altitude is the default for air tracks, speed for surface and sub tracks.

The range display includes additional red horizontal hairlines indicating the release ranges for typical weapons, e.g. Exocet, Harpoon, torpedo.

The scales for the different displays are <u>not</u> adjusted according to actual data. The altitude scale ranges from 0 - 40,000 ft, the speed scale from 0-1500 kts, and the range scale from 0 - 40 nautical miles. The time scale ranges from -12 min to 0 min in 3 min steps to facilitate nautical calculations (a craft travels 1/10 of its speed [kts] in nautical miles in 6 min)."

ANY QUESTIONS?

Further recommendations?

Comparison to Norms Window

Introduction Comparison to Norms Window. [Slide: Comparison to Norms option I]

"The Comparison to Norms Window is supposed to provide a quick comparison of features for one contact to features for exemplars of other contacts. There are three different options for this window.

For a two-dimensional feature comparison task, the critical variables - e.g. altitude and speed - and the respective ranges for different platform types can be displayed in a two-dimensional plot. Data from the contact of interest are represented as points in this two-dimensional space: actual data as the track symbol and historical data as a line. It can be seen with one glimpse at the display whether or not the track's data fit (or did fit at any time) into the ranges for a given platform type. The two-dimensional display allows irregular and interdependent specifications for the speed and altitude ranges, e.g. a speed range varying with altitude.

The lines surrounding range areas are colored in order to relate them to platform types of different threat levels."

[Slide: Comparison to Norms option II]

"The second option uses a three-level coding to provide information on whether a track's data fit into the specific data ranges or categories for a platform type. For a quick comparison of whether a track's data fit into several ranges, a three-level coding will suffice: either the track's data fits exactly into the respective range, or it is uncertain (e.g. fits within a certain deviation or is not interpretable), or it doesn't fit. The thresholds for the code assignment can be defined during mission preparation. The coded data are displayed in a two-dimensional matrix, with the variables in the rows and platform types in the columns. Columns are separated to facilitate the classification task. Exact data can be displayed upon request, e.g. click on the matrix cell of interest. Columns therefore have rounded corners to indicate that they are active screen elements.

There are three different ways to code the data's fit: colour coding, fill pattern coding and shape coding."

[Slide: Comparison to Norms option III]

"The option III display uses analog scales to facilitate the comparison task. Speed, altitude and descent/ascent rate are displayed. The respective value ranges for different platform types are displayed as bars on the respective scales. The track's actual data are displayed as a white line in the respective variable display. The numeric value of the variable is displayed under the scale title."

ANY QUESTIONS?

Selection procedure

Selection for option II. [Slide: Comparison to Norms option II]

"Let us first go back to option II.

Do you prefer the colour code, the fill pattern code or the shape code? Please rank-order the alternatives using the option cards."

Questionnaire. "Please check the appropriate symbols on the questionnaire to indicate your opinion about the listed arguments. Feel free to add any arguments we might have overlooked."

Decision option I vs. II vs. III. "Which one of the three options for the Comparison to Norms Window you just have seen do you prefer? Please rank-order the options using the file cards."

Further Recommendations?

Template Window

Introduction. [Slide: Template Window]

Purpose. "The Template Window assembles lower level features and compares them against reference values. It relates individual events and presents a hypothesis for the situation based on an integration of events."

Description. "You can select a hypothesis for display by clicking on the respective button. Hypothesis buttons are ordered from the left by decreasing likelihood. The farthest left button (representing the most plausible hypothesis) is the default.

According to the hypothesis selected for display, the expected behavior of a track typical for this hypothesis is displayed using bars representing the time range for the behavior to occur. The actual occurrence of a track's actions are matched to this template by displaying with a triangle and hash marks when the behavior (if so) actually occurred.

The track's expected actions (e.g. "approaching", "descending" etc.) are displayed in the order of their expected occurrence from top to bottom. The respective time ranges are arranged as bars on a time line. The length of the bar represents the expected time range for the behavior to occur. This template moves along the time scale to the left hand side. The actions taken by the track are displayed as inverted white triangles for discrete events or hashmarks for actions that continued for some time (like e.g. approaching). The length of the hashmarks indicate the duration of the event."

ANY QUESTIONS?

Further Recommendations?

Intent Hypothesis (SABER) Window

Introduction. [Slide: SABER Window]

Purpose. "The Intent Hypothesis Window provides lists of evidence to support various hypotheses. It also presents assumptions that are required to accept a given hypothesis."

Description. "The Intent Hypothesis Window displays evidence and assumptions to hold for three hypotheses at a time, the farthest left being the most plausible. More hypotheses can be displayed upon request (clicking the respective button).

Only supporting evidence for each hypothesis will be displayed. Counter-indicators will not be displayed, for they will disqualify the hypothesis or figure as supporting evidence under another hypothesis (being displayed there). However, the assumptions list will inform you about any assumptions to be made in order to accept the hypothesis.

Since there may be several pieces of information to be displayed in the evidence list, we have to decide whether to filter the list or not.

In the unfiltered list version -- option I -- you may have to scroll along the lists to look up the information you are interested in.

In the filtered list version, the evidence to be displayed is selected in order to differentiate between the offered hypotheses. Thus, evidence which is part of all three displayed hypotheses (e.g. "descending") will not be listed."

ANY QUESTIONS?

Further Recommendations?

Response Manager Window

Introduction. [Slide: RespMan Window option I, ROE table]

Purpose. "The Response Manager Window provides assistance in using preplanned responses."

Description. "There are two options for the Response Manager Window. In both options you can select between three general strategies, e.g. deconfliction, maximization of ownship safety and avoiding engagement.

Option I arranges preplanned actions both on a time and a range scale in a two-dimensional display. The responses figure as time/range rectangles moving to the left along the time scale, indicating time and range thresholds for the respective actions as well as the permitted delay to perform them. The 0 = now line will stay at the same place; the actual range of the track is displayed as a horizontal line, moving downward as the track approaches.

To avoid mutual covering of the response rectangles, only their upper and left border lines are displayed (thus forming "response angles"). There is still time to perform a given response, if the respective angle forms a closed rectangle with the 0 = now and the current range lines. These closed rectanges are highlighted in order to inform you what the system recommends you to do right now.

The size of the time/range angles for a given action as well as their spacing on the horizontal axis will depend on the speed of the track (since less time for action by range interval is left when the track is moving faster). The vertical spacing (= range spacing) of the planned actions will be unaffected by the track's speed."

[Slide: RespMan Window option II, ROE table]

"Since range and time axes are related this way, there is a certain redundancy in the option I display. Option II takes this into account, displaying only the time axis. In this option, the responses are arranged from top to bottom in the order they are recommended to be performed. Each action is integrated in a time bar arranged on the time axis to display the earliest and latest time to do it. The range thresholds are displayed in a table format.
The equal vertical spacing of action "bars" in the option II display gives the opportunity to add prompt/feedback buttons to the window. If it is time to start an action, a button saying "Do it" appears in the same line as the action bar. If the you click the button, the feedback "Done" is displayed at the same spot. If you do nothing and the time to take the action passes, the action specifier is highlighted, and the button turns into a "Done?" button (which can be clicked to a "Done" statement) for 1 minute. If within this time you still do not react, the system assumes you did not perform this action and displays the feedback "Missed"."

ANY QUESTIONS?

Questionnaire. "Please check the appropriate symbols on the questionnaire to indicate your opinion about the listed arguments. Feel free to add arguments we have overlooked."

Decision option I vs. II

"Which option do you prefer?"

Further recommendations?

Rules of Engagement

Introduction. [Slide: RespMan Window option I, integrated ROE]

"The rules of engagement (ROE) are displayed as a table inside the Response Manager Window, or alternatively integrated in the display. The latter means that e.g. the action "Engage" will not appear until the ROE criteria for engagement are met (i.e. warnings issued, track approaching, within 25 nm etc.).

Support for the ROE table comes from the view that ROE are not immediately related to course of action decisions, but, being highly important, they have to be visible at *any* time. It is a well-documented phenomenon that increasing stress leads to decreasing working memory capacity. Therefore, whether the user chooses to follow the recommended course of action in the Response Manager Window or not, he should always be able to look up his ROE.

Support for an integration of ROE in the recommended course of action comes from the view that a ROE table as a static display tends to be ignored when the user pays increasing attention to other parts of the display (which may be continuously changing). This is especially likely under high stress conditions, when the user tends to focus on immediately action-relevant cues. Integrating ROE in the course of action recommendations would thus ensure that the user pays attention to them, but only as long as he considers the recommended course of action at all."

ANY QUESTIONS?

Questionnaire. "Please check the appropriate symbols on the questionnaire to indicate your opinion about the listed arguments. Feel free to add arguments we have overlooked."

Decision option I vs. II. "Would you prefer the ROE to be displayed as an explicit table or to be integrated into the response plan?"

Further Recommendations?

Track Priority List

Introduction. [Slide: Track Priority List option I (tag)]

Purpose. "The Track Priority List provides an integrated picture: ID, intent, priority and why, next action to be taken (preplanned response), and the latest time to take the next action."

Description. "Each line of the Track Priority List includes the track symbol, track number button, (max.) three intent hypothesis buttons, the status of the track, the next action and the permitted delay for this action.

The "next action" displayed is selected in the following way: the Response Manager Window presents time and distance ranges within which a given action is recommended or, at least, reasonable. There will sometimes be more than one recommended action at a given time/range point. Of this set of actions, the first one listed will be displayed as "next action" to maintain the preplanned action order. However, the "permitted delay" will be determined by the action with the *least allowed time delay*. Note that this can be a different action. We assume here that the first action listed has to be done before the next action can begin.

We can alternatively display the track's tag - if it's tagged - or the track's bearing to facilitate its identification. Note that a tag would have to be entered manually (and blank space would be left if the track is not tagged)."

ANY QUESTIONS?

Questionnaire. "Please check the appropriate symbols on the questionnaire to indicate your opinion about the listed arguments. Feel free to add arguments we have overlooked."

Selection. "Do you prefer the tag or the bearing to be displayed in the list?"

Further Recommendations?

USEFULNESS RATING

"We are well aware of the fact that nobody - including us - is perfect. Neither is our DSS draft. The purpose of this interview is to take advantage of your operational knowledge in order to build a tactically <u>useful</u> support system.

Please tell us how useful you find the different windows by sorting these file cards on three piles: a very useful pile, a useful pile, and a useless pile. In the meantime, I will build a screen which includes your selected windows."

SIMULATION OF USE

[DEFTT scenario is unfrozen]

[Slide: full screen]

"How would you use the DSS in the running scenario?"

APPENDIX B: INTERVIEW DATA, RAW NOTES

Table B-1. Demographic data on participants.

Participant No.		Rank (CO/TAO)	Pos CO/TAO Watch	Months CO/TAO Commands	Number of MED/ PERGULF Deployments
1	LT	TAO	6	1	2
2	LCDR	TAO	50	2	1
3	CWO3	TAO	12	1	5
4	CDR	CO	36	4	5
5	LCDR	TAO	36	2	2
6	LT	TAO	12	1	2
7	CWO4	TAO	36	1	2
8	CDR	CO	100	4	4
9	LT	TAO	36	2	3
10	LCDR	TAO	40	2	3
11	CAPT	CO	42	1	8
12	CAPT	CO	27	4	4
13	LCDR	TAO	100	6	3
14	CAPT	CO	49	4	1
15	CDR	CO	80	4	1
16	CAPT	CO	72	3	2

BEFORE DSS INTRODUCTION

10 "Look first at altitude, profile, speed; operators take care of the rest. Other variables are range and whether on commercial route."

10 "Define a priority track by weapons release range, being inbound, not-commair altitude, going fast."

11 Def. of a positive ID: visual ID, NCTR and lack of mode 4 IFF if detected by the same aircraft or section.

13 Important information is: is he within launching range, is he pointing at me, does he respond to alerts, what IFF does he squawk?

14 Integration of information to determine intent and ID is crucial. Weak link in current systems.

14 CO doesn't do the ID job, but has to accept ID as given and to think what to do about it.

14 Altitude is a key factor in determining intent. Descending is important indicator for hostile intent.

14 The key watch station is the one that works on ID.

16 User must be able to hook on track for analysis from the geoplot. Sometimes you want to analyze a track that is not yet on the TPL. CO won't enter track numbers manually.

ALERTS WINDOW

Interference with other systems

7 ACDS gives own alerts. DSS alerts are additional to these.

10 Good: alerts are parallel, don't interfere with console use (as AEGIS alerts do: have to be quitted before console can be used; interrupt work and thinking process)

11 NTDS Alerts come at a rate of 10 per minute. Mostly ID conflicts and dual tracks. Too much to handle; alerts function usually turned off. DSS Alerts would be additional to that.

11 Situation awareness is (and has to be) focused on geoplot. Any additional system will distract the user from his principal task.

What to display

1 Is there a filter excluding commercial airliners from the alerts list?

5 Are alerts also triggered by ownship sensors?

4 Option II: **Rec.:** history list as default, to be deselected if the alerts list is desired

6 **Rec.:** option II: general history button as a toggle switch to select/deselect history window for *every selected* track

6 **Rec.:** more concrete alert messages, e.g. "turned on Cyrano 4 FC radar"

7 Alerts history will seldom be necessary. If a problem arises, there will seldom be time to go back through the history list. What is the diagnostic value of alerts history?

8 Matching of symbols with track numbers very good

8 Time zone characters not necessary: everything in combat is Zulu time

8 **Rec.:** display of air warning status, weapons status, weapons posture, because these affect interpretation of alerts.

10 **Rec**.: personally editable alerts tripwires. Tripwires would be different between CO and TAO.

11 Strongly dislikes any "bunch of alphanumerics".

11 A battle group commander is interested in: "why is this guy red?" ID and supporting evidence for ID are the important questions.

11 CO's need advice of the kind: "don't shoot that one!"

13 There are different alerts categories: tactical alerts, fuel shortage on intercept craft, navigation safety alerts etc. Which ones are handled/issued by the DSS?

13 **Rec.:** Reword "Cancel" to "Delete".

13 Prefers option I. If the user is stressed, click buttons draw attention away. The information in the history box can be gathered from elsewhere on the screen.

13 Tactical alerts come in the Aegis environment with a buzzer signal. Unbuzzered alerts can be overlooked. **Rec.:** add buzzer signal to important alerts.

13 **Rec.:** hierarchy of alerts: alerts messages should only be displayed if important or several tripwires are hit and ID and IFF indicate a threat. Alerts should also be integrate several tripwire informations. This would avoid 6 alerts per track.

14 Computer line items are not very helpful. Pictures are more suitable.

15 Alerts Window is a dumb list. The tactician is interested in relevant information: **Rec.:** the list should be filtered.

Tripwires

13 Is there a discrimination between ID's, i.e. U/AE. U/AF, U/UE? All require different tripwires.

13 Is there a hierarchy of alerts, i.e. a filter that lets only the most important information arrive at the user? E.g. a track can hit several tripwires at a time, but only the most important ones give an alert.

13 Who defines tripwires and message text?

13 Exocet range depends on altitude and shooter (air/surface platform). **Rec.:** tripwires have to be different for different platform types.

Capacity

1 6 lines may not be enough in the option I display

7 6 alerts are not enough

5 "crucial" rating at first statement (full information) means that this is an important point, which however is not accomplished by option 1!

7 Disagrees with statement "all info about every track at *any time"*: it's not at any time because 6 alerts are quickly scrolled down by new alerts

7 6 yellow alerts can dump one red alert.

9 **Rec.:** scroll bar to review more than 6 alerts in AL1. 6 alerts may be enough in most cases, but information should not be lost.

9 Option 2: a whole track can be bumped out of the window if more than 6 cause trouble

10 **Rec**.: Alerts recall function for option I.

15 6 alerts is the right number. More would be too difficult to handle. But alerts should not disappear completely.

15 Superseding of alerts is an advantage. It indicates the importance of the current (superseding) alert. Information is not lost, if the tripwires are set in an intelligent way.

16 6 is a finite number. Capacity may be a problem.

Integration

2 **Rec.**: click on track on geoplot brings up pop-up alerts history

2 2 trackballs are difficult to handle, user must grab the 2nd ball before he can use it

11 **Rec.:** Alerts function <u>must</u> be integrated in the geoplot.

15 Integration DEFTT/DSS would not necessary be an advantage: geoplot is graphical, DSS text. Thought processes would be mixed if the system was integrated.

Color code

4 Color code for air warnings is given by AW. Color assignment is ok to indicate importance, but not to be related to air warnings

8 Colors conflict with air warning status. **Rec.:** different way to communicate importance

12 Color code conflicting with air warning status is unimportant

13 Color code: all displayed alerts in the drawing are tactically significant, regardless of color. Overlooked the white alert under the colored ones.

14 Color code conflicts with air warning status. **Rec.:** use different way to communicate importance or different color set.

Organization

4 Option II: Recommends listing order by currency and level: most recent alert on highest level. Recency/importance dilemma is no problem as long as the logic to determine the alert to be displayed really can tell the most important alert

7 Active perceptual filtering of relevant information is happening anyway. This is no disadvantage of option I.

7 Option II: if 2 alerts about a track arrive at the same time, you won't be able to see the second unless you click "History".

8 If an alert is canceled, does a previously dropped one come up?

9 **Rec.:** more highlighting of alerts about selected track

10 The time sequence of alerts is unimportant. Alerts are "mentally" ordered by track and threat priority.

10 **Rec.** option II: would like top priority in top line. (accepts that TPL does exactly this)

13 **Rec.:** Track symbol at the end of the line, not at the beginning. Information in a line should be ordered by importance, and the symbol is not that important. The usual order is time-track # - rest.

13 Option II: what if a white alert comes after a red one? Does the color stay?

13 **Rec.:** hierarchy of alerts: alerts messages should only be displayed if important or several tripwires are hit and ID and IFF indicate a threat. Alerts should also be integrate several tripwire informations. This would avoid 6 alerts per track.

15 Prefers option II: would like to know first which track the alert is about, then look up the message.

15 Superseding of alerts is an advantage. It indicates the importance of the current (superseding) alert. Information is not lost, if the tripwires are set in an intelligent way.

16 **Rec.:** group option I per track.

User Dialog

7 Since DSS alerts are display alerts, there should be no action required to see them. On ACDS user has to sequence through alerts to see them all.

9 Prefers AL1 because less interaction

9 **Rec.:** scroll bar to review more than 6 alerts in AL1. 6 alerts may be enough in most cases, but information should not be lost.

10 **Rec**.: Alerts recall function for option I.

11 **Rec.:** move return to list button.

12 Good: Alerts Window doesn't overwhelm user

13 Prefers option I. If the user is stressed, click buttons draw attention away. The information in the history box can be gathered from elsewhere on the screen.

14 Who will be clicking on buttons? CO and TAO won't!

14 CO should never get immersed in computer operations.

15 Prefers to keep hands off as much as possible, but here it is good that action is necessary to see the history. It allows the decision maker to focus on COI.

16 **Rec.:** If a track has been canceled, it should return to the option II list if hooked on (on TPL, geoplot).

Use

10 Use: check alerts against own priorization.

11 Will use generally geoplot, actually the large screen display.

14 Are alerts supposed to be at the beginning of the detect to engage sequence?

14 Alerts window would be useful for other watchstations, but he as CO would not allow himself to have attention drawn away from the big picture by single-track information

14 Questionnaire: many items negligible because he wouldn't use the window.

15 Alerts II consists actually of two different windows. History display is a separate function from the alerts function.

General Criticism

11 Situation awareness is (and has to be) focused on geoplot. Any additional system will distract the user from his principal task.

11 Strongly dislikes any “bunch of alphanumerics”.

11 A battle group commander is interested in: “why is this guy red?” ID and supporting evidence for ID are the important questions.

11 Max 4-5s are available to handle a track in tough situations. If more time is available, there is no need for DSS

14 **Rec.:** rename window. Alerts is misleading. Alerts are indications that something bad will happen.

15 The system should do only what it can do at 100% accuracy.

15 Alerts II consists actually of two different windows. History display is a separate function from the alerts function.

TRACK PROFILE

Interference with other systems

4 Track history feature already exists in the AEGIS environment

What to Display

2 **Rec.:** Altitude vs. Range display with minimum ownship engagement ranges (“fire windows”)

4 **Rec.:** polar display for range history

11 Weapons release ranges are useful.

12 Would expect time 0 at the left hand side of the scale, but the scale is understandable if you are used to it

12 Combination of speed and altitude in one display is worth to investigate

13 Speed history is unimportant, missiles can be launched at almost any speed.

13 Range history is not needed.

13 **Rec.:** altitude vs. range display. This is the format most tactical manuals use. Profiles are always altitude over range, not over time. The range should be increasing to the right in order to be consistent with manuals. Radical descends on the profile are indicative of tactical actions.

13 **Rec.:** cross-reference alt/range display to threats, i.e. missile launch ranges (which depend on altitude).

14 **Rec.:** Altitude, Speed and range in one window. Altitude as a line (as is), speed and range as digital display boxes with + and - signs to indicate up/down trend. More detailed speed and range information is not needed, but it has to be there in the context of the altitude information.

14 Combined speed/altitude graph not better. Two scales would be necessary.

14 **Rec.:** Range can be more important than time to determine a profile. But to display weapons release ranges might be suggesting a hostile intent.

15 **Rec.:** Altitude over range display would be better. Range allows translation from the geoplot to the Track Profile. Users are trained to convert ranges and speed into time, not vice versa.

16 Displays based on range scale are better than time-based ones.

Scale Ranges

2 Track with altitude > 40.000ft can be a threat, **Rec.**: modifiable scale

4 The tactically important speed information is between 0 and 500 kts. Scale range thus too large.

6 **Rec.**: range scale should include 80nm for both air and surf/sub tracks

6 **Rec.**: increase time range for surf/sub tracks because they are slow. 30min will do

7 **Rec.**: 50kts speed scale range for ships. 100nm for range should be plenty.

8 Certain Soviet weapons (AS4, AS6) are 60-70Kft weapons, may be sold in gulf area. Indian Sunburns go Mach 2.4.

8 **Rec.**: log scale

10 Range: 40nm too late to make decisions. 80nm better

11 Scale range problem: high altitude weapons are very fast and thus easily detectable. Low altitude tracks are more a problem. Below 100 ft altitude is unreliable. Interesting tracks are low-medium altitude flyers.

11 Scale range for range depends on shooter.

12 **Rec.:** users should be able to define their own scale ranges

13 **Rec.:** 2 common scale ranges from 0-10 and 0-80Kft altitude, and 0-85 nm and 0-300 nm for range (numbers are examples for the order of magnitude)

14 12min is a long time. **Rec.:** make a time zoom function available with a narrower time line.

15 Constant scales are good. Window provides amplifying information without adding confusion.

15 Maximum resolution is not necessary. TPF is used to see the altitude trend; the actual value can be read from CRO.

Tactical Use

4 Descending CommAir slows down, attacking craft will speed up

8 Meaning of descending varies with range.

9 Would use range display first because range is (his) main priorization cue

10 TPF can backup CNI functions.

13 Would only use an altitude by range display.

13 Profiles exist only for missiles, not for platforms. Platform behavior gives no cue whether he can shoot.

14 Data should not be interpolated too much from a graph.

14 Would lock on window a big time in order to assess whether other information (from CIC team) makes sense.

16 Range display takes a moment to read.

Integration

2 **Rec.:** Altitude vs. Range display with minimum ownship engagement ranges ("fire windows")

11 TPF would be not bad if integrated in geoplot, i.e. in a 3-D display.

Migration

6 CN and TPF would be perfect assistance for AAWC, because they would take load from EW

User Dialog

8 **Rec.**: Range display: self-defined tripwires.

12 **Rec.:** users should be able to define their own scale ranges

General criticism

2 Tactical significance of range display is hard to interpret

3 Doesn't see the purpose of the window.

3 Operators may rely too much on the system.

11 Too much focusing on one guy. Overall tactical picture may be lost if user dives in one track's data.

COMPARISON TO NORMS WINDOW

Display Format

4 Option II: **Rec.:** to flip matrix and incline text on x axis, because platform type is the information of interest.

4 **Rec.:** option I and option II simultaneously.

12 Option II: relation to boundaries can be looked up by clicking on the cell of interest, so it's no problem that it is not contained in the display itself

14 **Rec.:** angle off platform specifiers by 30 degrees.

15 **Rec.:** CN2 would be good if added to CN1. Use CN1 as standard, CN2c if additional information is requested.

Variable set

2 Option II: Include maximum speed as separate variable. A maneuver might drop a fighter temporarily out of the fighter category when it becomes "too slow"

5 **Rec.**: Rename "Origin" to "Take off site".

5 **Rec.**: max. speed as additional variable for option II.

7 Option I: a 20min history tail will be enough

9 **Rec.**: CN2: cancel intel as variable. Hardly useful.

10 **Rec.:** CN2 rearrange variables according to personal preference. He prefers IFF at top, then alt, spd, EW, D/A rate.

11 **Rec.:** CN2: add variable NCTR / SARTIS "non-cooperating target recognition".

11 CN2 Time in air: how do we know? Origin is good.

11 IFF Mode 4 for ship-air response only reliable if <80 nm. AWACS' reliability range is about 350nm. But only USN has IFF 4.

12 **Rec.:** option I: also display IFF and EW information in text/table format

12 Option II: multiple parameters are a crucial advantage of this option. But history is missing.

12 Prefers option III to I because of more variables.

12 Option III: the three variables selected are the key ones.

14 Time in air: additional tanks under a fighter's wings are possible. Variable will not discriminate.

Platform type set

4 **Rec.**: rename "Fighter" to "military tactical".

6 **Rec.:** option II: include fighter, bomber, attack aircraft as additional platform categories

8 Drones are possible platform types.

9 **Rec.**: add bomber as platform type.

11 General: add platform type "missile" .

11 **Rec.:** Use commonly known abbreviations for platforms: VA/VF, HSL etc.

Tactical Use

4 Usefulness depends on proper definition of ranges

4 Option II: a column which is half yellow is as diagnostic as a full green column.

6 Adversary's intel (if it functions) can tell a fighter pilot how to fake a profile of a commercial. Variable ranges thus can be very large and undiagnostic

6 Option I's value comes from the parts that are redundant with TPF: display of altitude and speed including historical data.

6 CN2's EW information would be great if EW is busy

7 Option II information density may be a disadvantage, overwhelming the operator

8 Intuitive use is no factor because of months of training.

8 CN2 matrix is "what you do anyway". Certain schools teach exactly this way of processing information.

9 Prefers CN1 because it takes less time to look at.

10 Prefers CNI if stand-alone, prefers CN2 in combination with TPF.

10 CNI is not a backup for TPF. Would not look up a profile on CNI, but use it for classification only.

10 Prefers CN2 because CNI can be backupped by TPF.

10 CN2 "nice" as EW backup. Gives ability to measure team effectiveness, helps to get a better feeling about the team's decisions

11 CNI: There will be a big overlap of ranges.

11 CNI: adds nothing that is not intuitive anyway.

11 Prefers CN2 for "computational speed".

11 Used decision matrix before, but has been dropped. In training sessions it turned out that "good guys could be shot".

12 Option I redundancy with TPF is an advantage: user doesn't lock on one display

12 Option II non-obviousness is not a problem because of training

13 The only information differentiating really between commercial and military planes is visual ID and SARTIS NCTR. Being a bad guy means turning any EW off and create confusion in order not to be shot.

15 Norms are different for different areas of the world.

15 CN1 is easier to tune for local requirements and gives a feeling for possibilities and impossibilities. Recon craft flying consistently lower than 10 Kft can easily be integrated in the display.

15 CN2 is more driving to the best fit solution, which is not always reality.

15 CN2 variables are cumulative over time. May be better later in the process, but for new tracks it only displays speed and altitude.

15 CN2 doesn't give a feel why a platform type is excluded.

16 CN2 may be leading if there is no red in two columns but more yellow in one of them. Actually it could be both platform types.

Coding issues

4 Option II: likes color code, but dislikes pastel tones. **Rec.:** saturated colors.

6 Option II: Prefers color code because he's used to colors.

7 Option II: strong, saturated colors are subjectively better because of blurry eyes in bad CIC lighting conditions

8 CN 2 c: omissions look like a screen malfunction.

9 **Rec.:** CN1: threats should be red. Use saturated colors.

10 **Rec.:** change color set for CNI. More red/highlight for fighter range, because he's a threat. Color set according to priority in the order fighter>helo>recon.

13 **Rec.:** variation of option 2b: use smaller squares instead of omissions or empty squares for misfits to avoid confusion with missing data.

15 CN2c may be the quickest, but a is more familiar.

Migration

14 CN I not useful for CO, but for ID operators. It might be dangerous for the CO because it should not be the only tool to be used.

14 CN2 would be something for a watchstation operated by an enlisted man. Therefore a lot of training would be required.

General Criticism

7 Option II information density may be a disadvantage, overwhelming the operator

8 All CN options are useful, but prefers 2.

9 CN2b adds confusion to the screen.

11 CNI: There will be a big overlap of ranges.

11 CNI: adds nothing that is not intuitive anyway.

13 No CN window is really necessary. Would use it as an "afterthought only.

15 CN2 is "more useable and more dangerous". Pilots dont't always fly within specified parameters. CN2 makes more of a decision than it should. It is dangerous to exclude platform types!

15 CN3 extremely hard to use, "takes forever". Looking at different variables involves memory load.

TEMPLATE WINDOW

Display Format

13 **Rec.:** The movement of the templates has to be extrapolated between the 1 min updates. Otherwise the window will be very misleading.

14 **Rec.:** bottom line should be a range instead of a time scale. The main display is the geoplot which is not based on time, but range. Time scale is confusing: tactics are based on range; users are never trained to calculate times.

15 **Rec.:** A range scale would be better than the time scale (see arguments for range-based TPF).

15 The Template is a checklist. Would like an altitude over range display with activity areas. "Descend" and "Turn inbound" can then be seen immediately.

15 Coherency is crucial. A graphical display based on altitude and range can be compared to a nominal attack profile.

16 Graphical altitude/range template would be very good because much easier to absorb.

Tactical Use

10 The main difference between harass and attack is whether he carries weapons or not. Only information source for this is intel and visual ID.

11 "You would never go this deep for one track." [if ID unclear get a visual anyway]

12 It is good to provide a default hypothesis which is also the most likely one.

13 Expected behavior is strongly dependent on weapons. Nothing in the template means anything unless it is known whether he carries missiles or bombs.

13 Attack/Harass differentiation can be made with CPA: a close CPA is indicative of harassment. A true attacker in cold war would try not to look as an obvious attacker and to reach an optimum launch range with a wider CPA. He would not necessarily use ESM and fly a non-threatening profile, faint an air distress or even use a recon craft.

13 The first pass may be for harassment, but the second may be an attack.

13 The information differentiating best between attack and harass is intercepted communication, intel and weapons carried.

14 The template had to depend on ID. Because ID is hard to determine he would have no confidence in the system.

14 What happens with the Template display if ID changes?

14 Track Priority List tells intent and priority, so what is the need for the template window?

14 Would use the Template to look quickly whether or not triangle marks are there or not.

14 If a track doesn't fly an attack profile there is no interest in intent.

15 Coherency is crucial. A graphical display based on altitude and range can be compared to a nominal attack profile.

15 The Template is a checklist. Would like an altitude over range display with activity areas. "Descend" and "Turn inbound" can then be seen immediately.

16 What's the difference between an attack and a harass template? If there is not a good discrimination between hypotheses the template is meaningless!

16 Maybe attack vs. non-attack is the only question.

Coding/Wording issues

10 **Rec.:** reword "attack" to something less aggressive ("hostile" would conflict with ID). Unexperienced TAO trainees tend to be aggressive "beyond ROE".

14 **Rec.:** code likelihood of hypotheses by highlighting or color. Changing the position of the button is not enough.

14 **Rec.:** rename "attack" to "Attack Profile". "Attack" alone is leading.

General Criticism

3 Doesn't like window.

4 Doesn't like window.

11 "forget it"

11 "You would never go this deep for one track." [if ID unclear get a visual anyway]

14 The template had to depend on ID. Because ID is hard to determine he would have no confidence in the system.

14 There is a loss of confidence in the usefulness of the window because of the inconsistent and contradictory content of the example window.

SABER WINDOW

What to Display

1 Should show <u>all</u> information, not only differentiating info.

5 **Rec.:** toggle button to filter/unfilter list. Filter is good when user is familiar with the window's operation. In the beginning unfiltered version better.

7 Prefers filtered list: looking at additional info doesn't help verifying the system's suggestions

7 **Rec.:** selective filtering based on situation: different filters for different hypotheses. E.g.: accelerating during an attack looks different from accelerating because of an air distress

8 No strong feeling about preference

9 Prefers 1: system should not withhold information. Would like to know what is filtered.

9 **Rec.:** dim non-discriminating info (which would else be filtered).

10 Prefers unfiltered list: "questions unanswered may remain"

10 **Rec.:** add information about weapons.

11 **Rec.:** Add "No ID no IFF".

13 Attack/Harass differentiation can be made with CPA: a close CPA is indicative of harassment. A true attacker in cold war would try not to look as an obvious attacker and to reach an optimum launch range with a wider CPA. He would not necessarily use ESM and fly a non-threatening profile, faint an air distress or even use a recon craft.

13 The first pass may be for harassment, but the second may be an attack.

13 The information differentiating best between attack and harass is intercepted communication, intel and weapons carried.

13 Air distress: someone with communication problems would fly a triangle as an emergency signal.

13 In cold war an attack would probably look very un-threatening. One guy vs. a battle group would try to hide as long as possible. On the other hand, a terrorist could behave in any way. Behavior would be very culture-specific and differ from the Gulf area to North Korea. Military attacks would differ from terrorist attacks.

14 SABER information is only "good for academic reconstruction exercises" or computer people verifying the algorithm. For these purposes option I is better.

15 Decision maker has to know what is filtered. Otherwise too much training is required. **Rec.:** list common items separately on the right.

15 Not to display SABER forces the user to think by himself. **Rec.:** make operator form his own list.

16 SABER II better: common's don't contribute to the picture.

16 Doubt that three hypotheses would show up.

Use

6 SABER's use is that you can see all data the system holds about a track

11 Useful only if time. If time no need for it.

14 SABER information is only "good for academic reconstruction exercises" or computer people verifying the algorithm. For these purposes option I is better.

14 Wouldn't test assumptions on alternative hypotheses.

14 CO and TAO decide upon accepting the given ID, but they don't work on the ID problem.

15 Has already tried to use tactical tripwires.

15 **Rec.:** SABER tripwires must be editable by an ownship operator: there is a set of common actions for different threats. The question is, what is the critical information?

15 Not to display SABER forces the user to think by himself. **Rec.:** make operator form his own list.

Integration

2 Usefulness may be masked by wrong positioning. **Rec.:** pop-up window when a track is hooked on.

General Criticism

4 Would not trust the system. If the system works perfectly, it may be useful. If not, it is useless.

5 "The more you look at it the better you like it".

11 Boxes too small.

11 Useful only if time. If time no need for it.

11 ID classification is much more important than intent, because ROE are purely based on ID. Intent can be derived from ID.

11 "Would never believe what a machine tells on intent. Would not even want it."

11 Def. of a positive ID: visual ID, NCTR and lack of mode 4 IFF if detected by the same aircraft or section.

11 Intent reasoning turns attention away from ID (which is crucial).

11 If all relevant variables are included in the SABER evidence lists the screen would be too cluttered.

RESPONSE MANAGER

What to Display

2 **Rec.:** feedback "Done at ...time".

4 Option I: **Rec.:** color code to give feedback about done and missed actions. What has been done should be entered by someone else.

6 **Rec.:**ROE-based recommendations (e.g. "don't engage w/o warning") should be colored

9 **Rec.:** RM1: scroll the range scale in some reasonable way. Angles would not only move left, but also up with the range scale.

10 **Rec.:** finer time scale for fast attack fighters (which go 100nm/min).

12 Likes range scale, but actions should be decompressed.

12 Also likes option II, especially to list items and the status of the process

13 ROE and battle order are tied to ranges, not to time. **Rec.:** display a range scale instead of a time scale. Time scale goes into micromanagement and is not needed.

13 **Rec.:** include items "verify systems operability" and "inform crew". Report to senior happens after every event. Instead of "engage" put "request permission to engage". Add before 1st warning "req. visual or 3rd party ID".

13 RespMan is mostly needed to make sure that legally required actions take place. Thus, it should include only actions that “leave the ship”, i.e. warnings, illumination, reports etc. Most other actions (that stay within the ship) will be delegated anyway.

13 **Rec.:** reverse order of lock on and illuminate. Illuminating happens after warning. Action order has to be compatible with ROE statements.

14 **Rec.**: RM and Template in same format, but not in the same window.

14 Combination of range and time scale is very good.

14 **Rec.:** get rid of “report to senior”. This is not top priority.

14 Good: RM1 shows actual range of the target.

15 RM2 is a very effective time line list. Tabulated range is sufficient.

15 Prefers RM2 because easier to read.

16 RM depends on weapons. How many response sets are available?

16 Time scale is better than range scale for this purpose.

Use

7 Some ships use scripts as preplanned action lists which look much like option II

7 **Rec.:** time-tag actions and make them available for printout as an action report

7 option II click buttons are “necessary”. CO can tell at once what TAO has done

10 Used similar list (as RM2) for engagement procedures.

10 Use: like checklist. You can tell where you stand.

11 RM1 “not bad” if time and no more than 1-2 tracks to manage.

Integration

2 **Rec.:** Feedback about accomplished actions should come from confederate consoles.

5 **Rec.:** option II: cancel the buttons. Automate feedback.

6 RM II: buttons keep user busy. **Rec.:** cancel them. Automated feedback from effectors is necessary

11 Useful maybe as addition to 3D overview. But, where to put it?

User Dialog

4 Option I: **Rec.:** color code to give feedback about done and missed actions. What has been done should be entered by someone else.

5 **Rec.:** option II: cancel the buttons. Automate feedback.

6 RM II: buttons keep user busy. **Rec.:** cancel them. Automated feedback from effectors is necessary

7 Option II: user can forget to click although he did perform the action

7 **Rec.:** for option I collapse done actions to make more room available for the currently recommended actions

7 option II click buttons are "necessary". CO can tell at once what TAO has done

9 **Rec.:** RM2: skip buttons, have someone click them.

11 **Rec.:** Need better word for "deconflict". Can mean deconfliction of friendly weapons. Better: "continue ID"

11 RM2: who operates the click buttons?

11 Prefers II because better readable.

13 Three strategies are not needed. **Rec.:** The RM should be dedicated to legally required actions only.

15 7013 is one of 127 tracks and 37 warnings. Feedback function is only valuable if up to date. This is hard to be done. 6 actions in 3 minutes are 2 per minute. Consider three tracks: One feedback every 10 sec. Feedback thus is not going to be accurate!

Interference with other systems

11 Illumination is automated on Aegis.

General Criticism

9 RM2: continuous planning is a necessity. "No fallback into action planning" is a weakness!

11 RM leads user off the air picture.

ROE DISPLAY

What to Display

1 TAO's are trained to: "Don't let ROE become a checklist!"

5 Dislikes checkmarks: ROE are an aid/a recommendation, but not a concrete prompt to do things. **Rec.:** cancellation of checkmarks, option II disqualified!

6 ROE usually more complex than ours, "beautiful" if a ROE aid could be displayed on 1 screen

7 ROE may be platform-specific. If the ROE support system is case-sensitive, the small table may suffice

7 **Rec.:** combination of both options: do not display ROE table until engagement is an option (ROE then become central)

8 There may be several parallel courses of action. ROE table takes care of all

8 ROE table may not be ignored. Need for active integration in action plan is a positive (disagrees with questionnaire statements).

10 ROE box would be unimportant.

11 Prefers integrated ROE because "box has a tendency to be ignored".

11 Real ROE won't fit in this box.

12 Key points of ROE are needed dependent on threat. Prefers table.

12 Checkmarks are ok.

12 Whether ROE fit into the table depends on how smartly they are "boiled down"

13 Table adds to clutter on the screen. **Rec.:** clean up RM, put ROE in a small "remember" box at the bottom of the window. Usually there is no sequence in ROE.

13 What about two separate sets of ROE, e.g. for different countries? This is not uncommon.

14 Likes ROE box because it displays facts, while RM displays tasks.

User Dialog

5 User needs control over table contents.

Integration

6 ROE integration will require proper feedback about done/missed actions: "This will require interaction which may distract from a fast moving tactical situation.".

Use

4 ROE guidance is separate from single items on the action plan.

6 ROE display "nice tool for reminders", but applying ROE is a dynamic event. You may not engage even if the ROE rules have been met.

10 "Engage" bar depicts a time window and is not a prompt. The decision to engage is made way before the bar shows up

15 ROE and RM are related but separate. ROE are advice to ensure friendly forces safety and not to start unintended conflicts. TAO shall understand ROE, but follow RM course of action.

Checklist Dilemma

1 TAO's are trained to: "Don't let ROE become a checklist!"

7 ROE <u>are</u> kind of a checklist when it comes to engagement. They don't give a permission to shoot, but they must be met before shooting. But **Rec.:** get rid of checkmarks (they look like a permission to shoot).

8 CO's battle orders are written in numbers, ROE are not. Current ROE don't contain any numbers (e.g. range specifications).

8 ROE incompatible with digital thinking.

10 Anything leading to ROE-based decisions is almost thoroughly guided by checklists. So in fact it's a checklist decision.

13 ROE is a checklist. The TAO is controlled by the CO in order to prevent premature actions.

13 That ROE are met doesn't mean that the analysis of a track is done.

16 Often ROE are checked but there is some overriding thought in the CO's mind.

General Criticism

4 ROE usually are several pages, not just 4 items

6 "I currently would not trust a computer to tell me ROE has been met. I can see someone saying "I shot because the system said so (unlikely, but possible)."

11 Weapons status never higher than tight. In the gulf it was never higher than yellow, safe. CO has to be told to engage. ID is "the crucial thing", ROE are always tied to ID. Intent is immaterial.

16 ROE can be very complicated. Not sure if ROE can be reduced in any way to a small box.

TRACK PRIORITY LIST

What to Display

4 **Rec.:** also range display to identify track

5 Abbreviation "NOI" unfamiliar, but probably best for "not of interest"

6 **Rec.:** color use in TPL for items like "engage" ("biggest help for "snapshot glances" at the screen")

7 Bearing information is available elsewhere.

9 **Rec.:** bearing + range. They belong together (if bearing then range!).

9 **Rec.:** Tag: add some more characters.

10 Prefers bearing. **Rec.:** add range.

10 Would use track numbers as tags. So tags would be useless to display.

11 **Rec.:** get rid of intent buttons. The rest is much the same as the Aegis list.

11 Displaying last alerts message instead if intent buttons would be "not bad".

12 Red "IMMED" flag is good. **Rec.:** next action "engage" should also flash or be flagged

13 Bearing gives an identification cue before the track is tagged, which is important for fast tracks.

14 **Rec.:** Bearing and range display to help find the track. Tag would be useful at the right hand side of the list in order to tell "what he is".

14 **Rec.:** take out administrative actions like reporting. The scheduler should be about weapons and engagement only.

15 Prefers tag. The operator tags only priority tags, so they can easily be seen. The track number is only used as a tag if the track is wanted to be shot. For friendly forces the ship's name or plane's side number is used.

16 Can call up track number on the geoplot. Tag is sufficient, but bearing is better.

Integration

2 Neither tag nor bearing are really necessary. Automatic tagging would be "nice".

2 What happens if track numbers change (possible if link tracks are temporarily lost and pop up again as ownship tracks)?

4 **Rec.:** integration with Alerts Window

6 Relation track symbol/number: ex. a function on the DEFTT screen to highlight a track of a given number

10 Uses console functions to locate track.

11 Displaying last alerts message instead if intent buttons would be "not bad".

Use

8 Bearing ok: TAO knows where a track is, but CO maybe doesn't. Good for communication

11 Users look (and have to) at geoplot, not at TPL.

15 **Rec.:** use tripwires (Alerts Window) with smart hierarchy, pick priority tracks manually. Automatization is unrealistic because of unknown, widely varying weapons release ranges. The way to do it is: pick track for TPL, enter possible weapons, enter assumed intent, follow RM actions.

Definition of "Priority"

11 Aegis priority list is ordered by last opportunity to fire.

13 The priority order is tricky to define. It currently (Aegis) is defined largely by time on top. **Rec.:** Here it should be a combination of ID (hierarchy: hostile > UAE > UUE > UAF), time until track reaches engagement range, and last opportunity to engage.

14 **Rec.:** Priority order: track's weapons release range and profile.

15 Priority definition: how is the likely weapon specified? This is the crucial variable to determine priority! Never uses Aegis priority list because Aegis priority (last opportunity to engage) is useless in ambiguous scenarios.

15 **Rec.:** use tripwires (Alerts Window) with smart hierarchy, pick priority tracks manually. Automatization is unrealistic. The way to do it is: pick track for TPL, enter weapons, enter assumed intent, follow RM actions.

15 CO and TAO define TPL in teamwork, also using CIC team input. CO is asked for resources (weapons) or informed about a threat. Automatically derived assumptions are usually too negative.

16 Definition of priority: current NTU system can select the worst track. Window capacity iis large enough, so there will not be much of a difference between priority models. Latest time to launch is ok.

User Dialog

9 Prefers least possible manual entry.

Organization

General Criticism

10 CDS priority list was too slow, never up to date, which spoiled functionality. Calculation time is very important.

11 Users look (and have to) at geoplot, not at TPL.

11 People that look at alphanumerics are "dangerous to their own guys".

15 CO and TAO define TPL in teamwork, also using CIC team input. CO is asked for resources (weapons) or informed about a threat. Automatically derived assumptions are usually too negative.

FULL DSS SCREEN

5

Selection: AL2, CN1, SA1, RM2a, TPL1

Case of oil helo: select track from TPL. Look at profile-altitude.

Looks back and forth between Geoplot and TPL to see what's important.

Arrangement of windows "pretty good": first interest in alerts, TPF and CN (I). Considers Template and SABER if track's a threat.

"User has to get used to RespMan".

SABER useful to explain the captain why TAO has done what he has.

Time line in RM useful.

Main display for monitoring is still geoplot, until alert occurs.

Template and SABER belong together because both deal with intent reasoning.

Busy screen is no problem after getting used to it (2.5hrs looking at)

6

<u>Selection</u>: AL2, CN2a, SA1, RM2a, TPL1

If busy, he would use Alerts, TPL, TPF, CN II and keep responses and intent reasoning in mind.

ASTAB(?) provides much of TPL's functionality.

<u>Case of oil helo</u>: he would use SABER because the track is slow and intent reasoning matters. For a fast track, without time, he would use the 4 windows mentioned above.

Once ID is clear he would ignore the CN information. Alerts and TPL used as a backup for memory and situational awareness

overall "excellent"

8

<u>Selection</u>: AL2, CN1, SABER 1, RM2a, TPL2

Alerts, CN1, RM2a "very good".

TPL, SABER, TPF "useful"

Alerts, CN1, RM2a used most of the time.

<u>Rec.</u>: piping them on key CIC places without action functions.

Complexity no problem because of 6 months training and hours on watch.

Case of high speed inbound track: focus on Alerts and RM.

<u>Helo</u>: SABER for in-depth thought. CN maybe. No quick reaction required, so analysis windows can be used.

Helos are most problematic contacts in today's pers. gulf

Will DSS be a new watch station?

9

<u>Selection</u>: AL1, CN1, SABER1, RM1a, TPL2

<u>Color in AL attracts attention</u>: first look at Al.

Next drop down to TPL to recommended action.

<u>Next look at right part of the screen</u>: what is displayed to support that?

<u>Remark</u>: tries to verify the system's suggestions.

Do we reach agreement? Does the system support my own decisions?

<u>Case of high speed track</u>: may not look at DSS at all. Look at range, course, IFF and immediately automate CIWS. "DSS is nice to have, but I would not rely on it. It won't take the place of training and experience."

Case of Iranian P3: is he identified as Iranian? If yes: ignore TPF and CN. Think about response all the way: go down to RM. Like RM close to SABER: windows are linked to certain extend. Use RM as a cue for actions.

SABER: would use the evidence supporting the hypothesis he thought about (which is "Attack"). Experience says that P3's usually don't harass, but sometimes attack.

If an unexpected hypothesis would show up, he would be surprised and take a close look at analysis windows and SABER evidence.

Template not very much looked at; perhaps least useful window.

Most of the time going back and forth between AL and TPL.

10

Selection: AL2, CN2a, SA1, RM2b, TPL2

If close to an engagement he would not use the DSS.

Would scan the TPL to see "who to worry about next": not the top line, but below.

Then validate information using CN2,

look at profile, use both CN and SABER to validate ID checklist.

Then "march through" RM2b, checking with CIC team what has been done.

RM Feedback buttons are not clicked until action accomplishment (e.g. warning) has been reported. If busy he would not press the feedback buttons.

Uses TPL and RM to brief CO about tactical situation (TPL) first and process state (RM) next.

Makes no difference between slow helos and fast inbound tracks. Use of the screen is entirely guided by priority assessment. He would scan TPL continuously (besides Alerts).

Rec.: Intel info indicating hostile intent should be displayed more clearly. Currently it will be dug in SABER.

"Heavy EW user". Tries to get as much information as possible out of EW.

RM is used as a reporting guide and memory backup.

11

Selection: AL2, TPF, CN2c, TPL2, RM2b, no SABER and Template

Rec.: take logic behind DSS and add it to 3D-display logic.

Verifying the machine's suggestions is crucial; people tend to rely too much on machines

12

Selection: AL2, TPF, CN2a, SA1, RM2a, TPL2

Key window is TPL to establish priority.

Look then down to RM what the response plan is for this track.

Quick looks on Alerts list every once in a while.

The rest of the screen is just available as a visual reference (points at TPF)

CN2 is important because of multiple parameters.

Template and SABER give useful background information, why it is a priority track

<u>Case of high speed track</u>: would get an alert and go right down to the RM to look up responses. Uses the right part of the screen as background information.

<u>Case of low, slow helo</u>: determine intent using SABER. Question in mind; what is classifying with respect on intent?

13

<u>Selection</u>: Alerts I, CN2c, SABER II, RM II a, TPL II.

<u>Rec.</u>: Arrangement of windows according to importance. Upper left TPL, below Alerts, below CN2; on the right TPF, RM, below that Template and SABER. Reason for this: ROE are based on ID and relative threat. RM should thus come "after" CN.

Would monitor geoplot almost exclusively and use DSS for amplifying information. From geoplot can also be seen where other friendly forces are who can take care of a problem. Spatial relationships visible on the geoplot are very important.

If a track is threatening (inbound, unknown) he would use DSS or/and information from a friendly craft closer to the track.

First look would be at TPL in order to verify if the system puts the threat on the first priority line. Track Profile and CN information is hopefully provided by the CIC team. Would ask the team before consulting the DSS and balance/compare the informations. Crucial is, what is more up to date. Information quality has to be evaluated. Verbal reports may be based on more recent information. The DSS will have the last priority except for ROE support. Accuracy of DSS can't be guaranteed, but he is used to and comfortable with team's reports.

Would use the DSS more in the first 45 min of the watch in order to get more situational background.

What about silkworm threat on the beach?

Often several tracks have to be dealt with simultaneously. The DSS focuses much on one track.

Fast tracks are more important than slow ones unless these are at short range. But this is no rule, it depends on the region and politics.

DSS will probably be more useful in training. Experienced users are not guaranteed; the DSS may help there too.

<u>Rec.</u>: dedicate the whole screen to a few individually selected windows.

14

Selection: AI1, CN2a, SA1, RM2a, TPL2

Will focus on TPL and talk to TAO based on this to see whether there is agreement.

Rec.: swap positions of TPF and Alerts window. This is a personal preference. (likes TPF more than Alerts Window)

Rec.: consider more watchstation interaction. CO doesn't need some windows, other watchstations do.

Co should not manipulate keyboards and mice. He is the only one with the big picture.

15

Selection: AI2, CN1, SA1, RM2a, TPL1.

Rec.: Alerts should have the track symbol flicker on the geoplot because geoplot is standard monitoring screen.

If 7013 turns inbound it will trigger an alert.

Look first at CN1.

First question: what is his ID? Platform type is essential for ID, especially if ID is unknown.

Next questions: can he shoot at me? How, with which weapons?

Go back to Track Profile: what is his history?

Eventually pick him for TPL (if self-defined).

Look at RM for responses under worst-case assumption.

Then look at the rest, if there is a problem. SABER and Template will be ignored most of the time.

Template with an altitude over range picture would be very valuable.

Rec.: More than 3 tracks can not be handled. Display altitude/range template and RM for every one of them. Update selection by the track priority list. Implement CN and SABER as pop-up windows. 3 parallel displays are needed in order not to focus on one track only.

Complete ID is more important than platform types.

Rec.: tune CN2 to the ID problem: profile, country of origin, EW, IFF, ID maneuver, RTF profile, tactical air corridor.

To focus on intent presumes that ID is known.

16

Selection: AI2, CN1, SA2, RM2a, TPL2.

Would use TPF and CN for a quick look. Not sure about confidence in SABER and Template.

TPL migh be significant.

RM won't come up in most situations.

Woud use TPF and CN most of the time, and Alerts.

Initially there is no difference between high speed and low-slow tracks.

Hypotheses not looked at because of lack of confidence in them.

Concern: is the flicker on the alerts list enough to attract attention, if the user is busy looking at CN?

TPL is good to have, would check periodically to make sure that not focusing on the wrong track.

Al, TPF, CN, TPL, RM are handy.

Template: the assumption is that there are unique and identifiable contents for different hypotheses. An altitude/range graphical template would be more meaningful, but still TPF and CN are more useful.

Modified template (alt/range) would be a lot of information for a guy just walking in.

CN1 is very good as long as the boxes are correctly specified.

TPL is already available. Al, TPL, TPF are mostly display technology.

APPENDIX C: COMMENTED RECOMMENDATIONS

In the following section recommendations given by subjects are discussed in detail. Recommendations are loosely grouped by issues. The author's comments are printed in *italics*; the subject's code number is given with each individual recommendation. Refer to Table 1 for additional information on subjects. This listing of recommendations shows the wealth of ideas that have been risen by subjects who have abundant experience in the subject matter as well as in just using computerized consoles. The reader may find several of his/her own ideas discussed here in detail.

ALERTS WINDOW

4 Option II: **Rec.:** history list as default, to be deselected if the alerts list is desired

6 **Rec.:** option II: general history button as a toggle switch to select/deselect history window for every selected track.

A general toggle switch would spare one click per track for a "history user" in order to see the history window. "Not-usually-history users" had to click twice in order to see the history, as they had to currently. The problem is that a default history window in the current window arrangement could easily cover the top lines of the Track Priority List for a longer period of time. It had also to be deselected if the user wanted to look up information on a less recent alert of a different track, whose track number button would be covered by the pop-up box. Since the main function of the Alerts Window is to attract attention on a problem track, the history function (rather supporting analysis processes) should not be a general switch in the experimental DSS.

6 **Rec.:** more concrete alert messages, e.g. "turned on Cyrano 4 FC radar".

This would lead to a mix-up of alerting and analysis functions.

8 **Rec.:** display of air warning status, weapons status, weapons posture, because these affect interpretation of alerts.

This is clearly not a DSS function. In an applied system this display would not be on the screen, but somewhere in the room.

8 Colors conflict with air warning status. **Rec.:** different way to communicate importance.

14 Color code conflicts with air warning status. **Rec.:** use different way to communicate importance or different color set.

SMEs consider this not to be a problem, unless the track density is very high.

9 **Rec.:** more highlighting of alerts about selected track

Track number buttons are already highlighted. Highlighting the whole message line is possible. However, this would support an analysis function the Alerts Window is not intended to support.

9 **Rec.:** scroll bar to review more than 6 alerts in AL1. 6 alerts may be enough in most cases, but information should not be lost.

10 **Rec.**: Alerts recall function for option I.

This is a useful extension of option I which solves the capacity problem. Actually a scroll button should be used which blinks if more than 6 alerts run in during a 6s update period in order to tell the user that something new is hidden by the lower window border (a standard scroll bar can not blink and is too small for quick operation). If the button is used to scroll down, a second button to scroll up again should appear at the top of the list..

10 **Rec.** option II: would like top priority in top line (accepts that TPL does exactly this).

The Alerts Window is not a management window, but intended to direct attention on tactically significant, recent information. It thus has to be ordered by alert recency.

10 **Rec.**: personally editable alerts tripwires. Tripwires would be different between CO and TAO.

This is surely necessary in an applied system. In the experimental system it would spoil experimental standardization.

11 **Rec.:** Alerts function <u>must</u> be integrated in the geoplot.

This is true for any applied DSS solution. In the experimental DSS functions have to be separate.

11 **Rec.:** move return to list button.

The participant had no better idea where to move it. However, if the history window is by mistake selected for a wrong track, the cursor will still be close to the "history" button field. In order to reduce necessary trackball movements, the "return to list" button should be close to this area. So the participant was right.

13 **Rec.:** Track symbol at the end of the line, not at the beginning. Information in a line should be ordered by importance, and the symbol is not that important. The usual order is time-track number - rest.

Track symbol and number are related by the convention to always display them together. Furthermore, the symbol indicates ID which is obviously very important.

13 **Rec.:** Reword "Cancel" to "Delete".

O.k.

13 Exocet range depends on altitude and shooter (air/surface platform). **Rec.:** tripwires have to be different for different platform types.

Recommendation is forwarded to SME's to be considered during knowledge base building.

13 Tactical alerts come in the Aegis environment with a buzzer signal. Unbuzzered alerts can be overlooked. **Rec.:** add buzzer signal to important alerts.

In DEFTT this would be the only buzzer. To attract attention, the flickering of new alerts messages should suffice, if no other flickering elements are on the screen (Wickens, 1984).

13 **Rec.:** hierarchy of alerts: alerts messages should only be displayed if important or several tripwires are hit and ID and IFF indicate a threat. Alerts should also be integrate several tripwire informations. This would avoid 6 alerts per track.

Recommendation is forwarded to SME's to be considered during knowledge base building.

15 Alerts Window is a dumb list. The tactician is interested in relevant information: **Rec.:** the list should be filtered.

The list is "filtered" by the specification of the tripwires.

14 **Rec.:** rename window. Alerts is misleading. Alerts are indications that something bad will happen.

This is intended.

16 **Rec.:** group option I per track.

This would lead to unavoidable confusion and conflicts between alert recency and track number. Simultaneous alerts are grouped by track number.

16 **Rec.:** If a track has been canceled, it should return to the option II list if hooked on (on TPL, geoplot).

Very good!

TRACK PROFILE

2 **Rec.:** Altitude vs. Range display with minimum ownship engagement ranges ("fire windows").

The Track Profile is intended to be an analysis window and not a management window. In an applied system ownship activity information can usefully be integrated to analysis information. In the experimental DSS functions are deliberately separate.

4 **Rec.:** polar display for range history.

This would include bearing history, similar to a trajectory function on the geoplot. Since all data displayed are relative to ownship, this is a very interesting option for surface tracks (intercept course can easily be detected).

6 **Rec**.: range scale should include 80nm for both air and surf/sub tracks.

The range scale will be 100nm.

6 **Rec**.: increase time range for surf/sub tracks because they are slow. 30min will do

O.k.

7 **Rec**.: 50kts speed scale range for ships. 100nm for range should be plenty.

O.k.

8 **Rec.**: log scale

Log scales are hard to interpret intuitively. Changes in the variables are impossible to asses immediately. Since change assessment support is the main function of the window, log scales are disqualified.

8 **Rec.**: Range display: self-defined tripwires.

This feature is already implemented in the Alerts Window and in the DEFTT geoplot (self-defined range circles can be entered around every object).

12 **Rec.:** users should be able to define their own scale ranges.

Possible solution for the problem of scale ranges. However, in the experimental DSS it would conflict with experimental standardization.

13 **Rec.:** altitude vs. range display. This is the format most tactical manuals use. Profiles are always altitude over range, not over time. The range should be increasing to the right in order to be consistent with manuals. Radical descends on the profile are indicative of tactical actions.

14 **Rec.:** Range can be more important than time to determine a profile. But to display weapons release ranges might be suggesting a hostile intent.

15 **Rec.:** Altitude over range display would be better. Range allows translation from the geoplot to the Track Profile. Users are trained to convert ranges and speed into time, not vice versa.

See discussion in main text.

13 **Rec.:** 2 common scale ranges from 0-10 and 0-80Kft altitude, and 0-85 nm and 0-300 nm for range (numbers are examples for the order of magnitude).

This would result to 4 alternative scale configurations which will spoil the intended consistent mapping.

13 **Rec.:** cross-reference alt/range display to threats, i.e. missile launch ranges (which depend on altitude).

This would allow cognitive operations which are not related to situation recognition by plain altitude and range history. Track behavior is put into a context different from the one first intended. This additional information could be used in a version of the Template Window.

14 **Rec.:** Altitude, Speed and range in one window. Altitude as a line (as is), speed and range as digital display boxes with + and - signs to indicate up/down trend. More detailed speed and range information is not needed, but it has to be there in the context of the altitude information.

Digital display boxes for the variables not included in the selected graph can be added to the window.

14 12min is a long time. **Rec.:** make a time zoom function available with a narrower time line.

The time scale can be adjusted to different requirements for surface and air tracks. Interactive functions should only be used if unavoidable.

COMPARISON TO NORMS WINDOW

4 Option II: **Rec.:** flip matrix and incline text on x axis, because platform type is the information of interest.

O.k. The participant's argumentation is correct.

4 **Rec.:** option I and option II simultaneously.

15 **Rec.:** CN2 would be good if added to CN1. Use CN1 as standard, CN2c if additional information is requested.

Incompatible with the experimental plan. Possible extension in side experiments.

5 **Rec.**: Rename "Origin" to "Take off site".

Take off site and origin are different things and can be separate variables in the option II display.

5 **Rec.**: max. speed as additional variable for option II.

This adds "memory" to the display which is certainly useful to determine platform type.

9 **Rec.**: CN2: cancel Intel as variable. Hardly useful.

It is: weapon's role can be recon and attack for the same aircraft type.

10 **Rec.:** CN2: rearrange variables according to personal preference. He prefers IFF at top, then alt, spd, EW, D/A rate.

The arrangement of variables is defined by the subject matter experts. Considering personal preferences and/or situational requirements is surely necessary in an applied system, but incompatible with experimental standardization.

11 **Rec.:** CN2: add variable NCTR / SARTIS "non-cooperating target recognition".

NCTR is not available in the experimental environment. Would certainly be useful in an applied system.

4 **Rec.**: rename "Fighter" to "military tactical".

O.k.

6 **Rec.:** option II: include fighter, bomber, attack aircraft as additional platform categories.

9 **Rec.**: add bomber as platform type.

O.k.

11 **Rec.:** Use commonly known abbreviations for platforms: VA/VF, HSL etc.

Currently unabbreviated platform names fit into the table.

4 Option II: likes color code, but dislikes pastel tones. **Rec.:** saturated colors.

The color set has to be defined considering perceptual issues in order to maximize the detectability of platform type columns without misfitting data. Saturated colors can interfere with the suggestion that green and yellow matrix cells both indicate compatibility with a hypothesis. In addition saturated colors would suggest a warning status.

9 **Rec.:** CN1: threats should be red. Use saturated colors.

10 **Rec.:** change color set for CN1. More red/highlight for fighter range, because he's a threat. Define color set according to priority in the order fighter>helo>recon.

Saturated red, yellow and green should be reserved for a warning context. Saturated red color should only appear on the screen if immediate attention is required. If saturated red would be used here, it would appear constantly on the screen, no matter what the track's data are.

12 **Rec.:** option III: also display IFF and EW information in text/table format

O.k.

13 **Rec.:** variation of option 2b: use smaller squares instead of omissions or empty squares for misfits to avoid confusion with missing data.

O.k.

14 **Rec.:** angle off platform specifiers by 30 degrees.

This does not increase readability on a CRT display.

TEMPLATE WINDOW

10 **Rec.:** reword "attack" to something less aggressive. Unexperienced TAO trainees tend to be aggressive "beyond ROE".

14 **Rec.:** rename "attack" to "Attack Profile". "Attack" alone is leading.

Users are assumed to be highly trained and experienced.

13 **Rec.:** The movement of the templates has to be extrapolated between the 1 min updates. Otherwise the window will be very misleading.

O.k. I agree.

14 **Rec.:** bottom line should be a range instead of a time scale. The main display is the geoplot which is not based on time, but range. Time scale is confusing: tactics are based on range; users are never trained to calculate times.

Important point. Time is a crucial factor in terms of the task, but obviously experienced users are well trained in converting ranges and speeds into a time "feeling". As long as we assume highly experienced users there would be no use in deriving time information from

range and speed. RM and Template displays thus can be extremely simplified because no moving bars are required, but just a speed leader.

14 **Rec.:** code likelihood of hypotheses by highlighting or color. Changing the position of the button is not enough.

Disagree. The position of buttons is intuitive and does not conflict with warning color codes.

SABER WINDOW

2 Usefulness may be masked by wrong positioning. **Rec.:** pop-up window when a track is hooked on.

Surely a very useful suggestion for an applied system, but incompatible with experimental standardization.

5 **Rec.:** toggle button to filter/unfilter list. Filter is good when user is familiar with the window's operation. In the beginning unfiltered version better.

Surely a very useful suggestion for an applied system, but incompatible with experimental standardization.

7 **Rec.:** selective filtering based on situation: different filters for different hypotheses. E.g.: accelerating during an attack looks different from accelerating because of an air distress.

10 **Rec.:** add information about weapons.

11 **Rec.:** Add "No ID no IFF".

Recommendations are forwarded to SM's to be considered during knowledge base building.

9 **Rec.:** dim non-discriminating info (which would else be filtered).

Helpful suggestion to improve the usability of option I.

15 Decision maker has to know what is filtered. Otherwise too much training is required. **Rec.:** list common items separately on the right.

It is easier to dim out common items. To list them in a separate place may be confusing.

15 Not to display SABER forces the user to think by himself. **Rec.:** make operator form his own list.

Probably too time-consuming. The DSS support consists in automatizing the evidence list formation.

15 **Rec.:** SABER tripwires must be editable by an ownship operator: there is a set of common actions for different threats. The question is, what is the critical information?

Definitely true for an applied DSS. For the sake of experimental standardization we use fixed SABER tripwires.

RESPONSE MANAGER

2 **Rec.:** feedback "Done at ...time".

7 **Rec.:** time-tag actions and make them available for printout as an action report

Useful suggestion for an applied system and for experimental data gathering.

2 **Rec.:** Feedback about accomplished actions should come from confederate consoles.

The experimental DSS will not be integrated with the confederate consoles.

4 Option I: **Rec.:** color code to give feedback about done and missed actions.

Feedback is required, but color should not be used because of color inflation.

4 What has been done should be entered by someone else.

9 **Rec.:** RM2: skip buttons, have someone click them.

O.k.

6 **Rec.:**ROE-based recommendations (e.g. "don't engage w/o warning") should be colored.

RM is intended not to include negligible actions, thus a common color can be chosen.

9 **Rec.:** RM1: scroll the range scale in some reasonable way. Angles would not only move left, but also up with the range scale.

User would have to read the scale in order to determine current range.

10 **Rec.:** finer time scale for fast attack fighters (which go 100nm/min).

O.k. Time scale has to be adjusted for air and surface tracks separately.

5 **Rec.:** option II: cancel the buttons. Automate feedback.

6 RM II: buttons keep user busy. **Rec.:** cancel them. Automated feedback from effectors is necessary

O.k. However, the experimental DSS can currently not be integrated with the confederate consoles.

7 **Rec.:** for option I collapse done actions to make more room available for the currently recommended actions.

Spoils the secondary use of the window to communicate state of the process. Previous actions would no no longer be displayed within time/range context.

11 **Rec.:** Need better word for "deconflict". Can mean deconfliction of friendly weapons. Better: "continue ID"

Deconfliction is an accepted fleet term for determining ID and issuing warnings.

13 ROE and battle order are tied to ranges, not to time. **Rec.:** display a range scale instead of a time scale. Time scale goes into micromanagement and is not needed.

Using a range scale leads to a conflict of scale direction (increasing ranges to the right) and reading direction (left to right). The scale has to be chosen so itis consistent with the Template Window.

13 **Rec.:** include items "verify systems operability" and "inform crew". Report to senior happens after every event. Instead of "engage" put "request permission to engage". Add before 1st warning "req. visual or 3rd party ID".

13 **Rec.:** RespMan is mostly needed to make sure that legally required actions take place. Thus, it should include only actions that "leave the ship", i.e. warnings, illumination, reports etc. Most other actions (that stay within the ship) will be delegated anyway.

13 **Rec.:** reverse order of lock on and illuminate. Illuminating happens after warning. Action order has to be compatible with ROE statements.

13 Three strategies are not needed. **Rec.:** The RM should be dedicated to legally required actions only.

14 **Rec.:** get rid of "report to senior". This is not top priority.

Recommendations are forwarded to SME's to be considered during knowledge base building.

14 **Rec.**: RM and Template in same format, but not in the same window.

O.k.

ROE DISPLAY

5 Dislikes checkmarks: ROE are an aid/a recommendation, but not a concrete prompt to do things. **Rec.:** cancellation of checkmarks, option II disqualified!

7 ROE <u>are</u> kind of a checklist when it comes to engagement. They don't give a permission to shoot, but they must be met before shooting. But **Rec.:** get rid of checkmarks (they look like a permission to shoot).

O.k. However, tracking ROE conditions met is necessary for a context-sensitive ROE support system.

7 **Rec.:** combination of both options: do not display ROE table until engagement is an option (ROE then become central).

13 Table adds to clutter on the screen. **Rec.:** clean up RM, put ROE in a small "remember" box at the bottom of the window. Usually there is no sequence in ROE.

To discuss context-sensitive ROE support is beyond the scope of this report. In the experimental environment, ROE can be displayed continuously.

TRACK PRIORITY LIST

4 **Rec.:** also range display to identify track

9 **Rec.:** bearing + range. They belong together (if bearing then range!).

10 Prefers bearing. **Rec.:** add range.

14 **Rec.:** Bearing and range display to help find the track. Tag would be useful at the right hand side of the list in order to tell "what he is".

O.k.Tag can be added if space is available.

4 **Rec.:** integration with Alerts Window.

Useful suggestion for an applied system. However, windows then are no longer functionally separate.

6 **Rec.:** color use in TPL for items like "engage" ("biggest help for "snapshot glances" at the screen")

DSS would be very leading then. Danger of color inflation.

9 **Rec.:** Tag: add some more characters.

Tags can be 8 characters long. The 3 character tag was intended to be an example.

11 **Rec.:** get rid of intent buttons. The rest is much the same as the Aegis list.

Intent buttons are integral part of the TADMUS DSS.

12 Red "IMMED" flag is good. **Rec.:** next action "engage" should also flash or be flagged.

The DSS is intended to support situation assessment, not COA decisions. Additionally, several subjects reported that unexperienced TAO's tend to engage too early. This would be even more likely if an action like "engage" is highlighted.

13 The priority order is tricky to define. It currently (Aegis) is defined largely by time on top. **Rec.:** Here it should be a combination of ID (hierarchy: hostile > UAE > UUE > UAF), time until track reaches (his) engagement range, and last (ownship) opportunity to engage.

14 **Rec.:** Priority order: track's weapons release range and profile.

Recommendations are forwarded to SME's to be considered during knowledge base building.

14 **Rec.:** take out administrative actions like reporting. The scheduler should be about weapons and engagement only.

The TADMUS Track Priority List is intended to support in ambiguous situations in low-intensity conflicts, i.e. way before engagement.

15 **Rec.:** use tripwires (Alerts Window) with smart hierarchy, pick priority tracks manually. Automatization is unrealistic because of unknown, widely varying weapons release ranges. The way to do it is: pick track for TPL, enter possible weapons, enter assumed intent, follow RM actions.

It has to be checked whether automatization can be done. If not, the suggested user dialog has to be implemented and refined.

FULL SCREEN

8 Alerts, CN1, RM2a used most of the time. **Rec.:** piping them on key CIC places without action functions.

Surely a useful suggestion for an applied system, but incompatible with experimental setup. DSS is designed for CO and TAO only.

11 **Rec.:** take logic behind DSS and add it to 3D-display logic.

Surely a useful suggestion for an applied system, but incompatible with current experimental plan.

13 **Rec.:** Arrangement of windows according to importance. Upper left TPL, under TPL Alerts, then CN2; on the right TPF, RM, below that Template and SABER. Reason for this: ROE are based on ID and relative threat. RM should thus come "after" CN.

RM currently comes "after" CN when the windows are used in clockwise sequence. RM should also "come after" the intent reasoning windows Template and SABER.

13 **Rec.:** dedicate the whole screen to a few individually selected windows.

This is intended after the objective evaluation of all windows.

14 **Rec.:** swap positions of TPF and Alerts window. This is a personal preference. (likes TPF more than Alerts Window)

The Alerts window is correctly placed close to the geoplot, but not in optimum display space. It has to attract attention from any part of the DEFTT-DSS system, but it is not looked at for a longer time.

14 **Rec.:** consider more watchstation interaction. CO doesn't need some windows, other watchstations do.

Surely a useful suggestion for an applied system, but incompatible with experimental plan. DSS is designed for CO and TAO only.

15 **Rec.:** Alerts should have the track symbol flicker on the geoplot because geoplot is standard monitoring screen.

This would be the way to attract attention in an integrated system.

15 **Rec.:** More than 3 tracks can not be handled. Display altitude/range template and RM for every one of them. Update selection by the track priority list. Implement CN and SABER as pop-up windows. 3 parallel displays are needed in order not to focus on one track only.

This is a completely different display concept which is worth to investigate after the current approach (functions and displays have to be integrated and shaken down to few displays).

15 **Rec.:** tune CN2 to the ID problem: profile, country of origin, EW, IFF, ID maneuver, RTF profile, tactical air corridor.

Note that the whole DSS's purpose is to support ID.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 1995	Final

4. TITLE AND SUBTITLE

SUBJECTIVE EVALUATION OF HUMAN-COMPUTER INTERFACE OPTIONS FOR A TACTICAL DECISION SUPPORT SYSTEM

5. FUNDING NUMBERS

PE: 0602233N
AN: DN300090

6. AUTHOR(S)

Bernard K. Rummel, German Naval Medical Institute

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Command, Control and Ocean Surveillance Center (NCCOSC)
RDT&E Division
San Diego, CA 92152–5001

8. PERFORMING ORGANIZATION REPORT NUMBER

TR 1698

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Tactical Decision Making Under Stress (TADMUS) personnel performed a subjective evaluation study to investigate human-computer interface (HCI) design options for an experimental decision support system (DSS). This report presents the study's results. The DSS contains seven windows: Alerts, Track Profile, Comparison to Norms, Template, SABER (Situation Assessment by Explanation-Based Reasoning), Response Manager, and Track Priority List. The study examined 19 individual design options for these windows.

Seven shipboard-qualified commanding officers and nine tactical action officers participated in the study. All were experienced in littoral warfare on Aegis or New Threat Upgrade (NTU) ships. The Decision Making Evaluation Facility for Tactical Teams (DEFTT) at NRaD collected the data. Researchers collected four data types: interview data, questionnaire data, preferences for individual options, and usefulness ratings.

The study's data will enhance subjective usefulness in the present task domain, and tactical decision making in ambiguous, littoral scenarios. It will also help to achieve maximum experimental control in future objective evaluation experiments. Researchers used user interview data, questionnaire data, and preferences to determine window design principles. They also used these data to determine the range of possible user behavior. The strong inter-individual variability in all three data sets is remarkable.

14. SUBJECT TERMS

human-computer interface (HCI)
Tactical Decision Making Under Stress (TADMUS) program
decision support system (DSS)
Decision Making Evaluation Facility for Tactical Teams (DEFTT)

15. NUMBER OF PAGES

120

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAME AS REPORT

UNCLASSIFIED

21a. NAME OF RESPONSIBLE INDIVIDUAL	21b. TELEPHONE *(include Area Code)*	21c. OFFICE SYMBOL
J. G. Morrison	(619) 553–9070	Code 44201

INITIAL DISTRIBUTION

Code 0012	Patent Counsel	(1)
Code 0271	Archive/Stock	(6)
Code 0274	Library	(2)
Code 40	R. C. Kolb	(1)
Code 44	J. Grossman	(1)
Code 44210	J. G. Morrison	(10)

Defense Technical Information Center
Fort Belvoir, VA 22060–6218 (4)

NCCOSC Washington Liaison Office
Washington, DC 20363–5100

Center for Naval Analyses
Alexandria, VA 22302–0268

Navy Acquisition, Research and Development
Information Center (NARDIC)
Arlington, VA 22244–5114

GIDEP Operations Center
Corona, CA 91718–8000

Office of Naval Research
Arlington, VA 22217–5000 (4)

Tactical Training Group, Pacific
San Diego, CA 92147–5087 (2)

Chief of Naval Operations
Washington, DC 20350–2000

Naval Sea Systems Command
Arlington, VA 22242–5165

Naval Surface Warfare Center
Dahlgren, VA 22448–5000

Naval Air Warfare Center Training Systems
Division
Orlando, FL 32826–3224

AEGIS Training Center
Dahlgren, VA 22448–5190

Armstrong Laboratory
Brooks AFB, TX 78235–5352

U.S. Army Research Institute
FPO AE 09499–1500

American Psychological Association
Washington, DC 20002–4242

ORINCON Corporation
San Diego, CA 92122